

## **ZONE 3 NECROSIS ASSOCIATED MARKERS AND METHOD OF USE THEREOF**

### **RELATED APPLICATION**

5           This application claims priority to U.S. Serial No. 60/410,763, filed September 13, 2002.

The contents of this application is incorporated herein by reference in its entirety.

### **FIELD OF THE INVENTION**

The invention relates generally to the identification of genetic markers associated with toxicity.

10

### **BACKGROUND OF THE INVENTION**

Necrotic cell death is a common response of the liver to hepatotoxic agents and represents an irreversible form of damage to individual liver cells. While the mechanisms by which hepatotoxic agents lead to necrosis remain to be fully understood, progress has been made in understanding the biochemical pathways involved. Necrotic cell death occurs when a chemical  
15 or it's metabolite react with critical cellular systems resulting in ion dysregulation, mitochondrial dysfunction and oxidative stress [1-3]. The acute disruption of these normal cellular events in effect leads to ATP depletion. This loss of energy within the cell distinguishes necrosis from the other classification of cell death known as apoptosis. When cells undergo apoptosis or programmed cell death, the cell requires an energy level capable of triggering special metabolic,  
20 signal transduction and gene regulation pathways that systematically shut down the cell. Necrosis occurs when the ATP levels falls below the threshold required for these processes and the cell is driven into a passive state of cellular chaos that culminates in cell death [4]. Thus, although these two forms of cell death are distinct, they can share initiating pathways depending on the how sharply the ATP levels decline. Such can be seen with the induction of the  
25 mitochondrial permeability transition (MPT), a mechanism that causes mitochondrial failure.

The MPT will lead to necrosis if ATP is depleted or apoptosis if there are sufficient amounts available to initiate a caspase cascade [5].

Chemical insult that produces necrosis of the liver can be either nonzonal or zonal. Zonal necrosis is separated into zones 1, 2 and 3 based on the region of the lobule affected. Different hepatotoxic agents preferentially target specific zones [6]. This research project report specifically deals with those agents that produced zone 3 or centrilobular necrosis. This is the most commonly affected area of the liver for hepatotoxic agents producing zonal necrosis. Zones 1-3 are distinguishable in terms of blood flow, oxygen content, bile flow and ratio of intoxication versus detoxification pathways. Factors such as these explain the specificity of hepatotoxic agents for particular zones. Acetaminophen and carbon tetrachloride (CCl<sub>4</sub>) are examples of agents that produce mainly zone 3 necrosis once they are converted to reactive metabolites. This can be attributed to the high degree of regional organization of agent specific cytochrome P450's within the liver [7]. The supply of oxygen available to the cell has also been shown to be a factor for zone 3 necrosis producing agents. Zone 3 is the region of the liver that is furthest from the arterial blood supply receiving the least supply of oxygen. When CCl<sub>4</sub> is metabolized to its reactive metabolite CCl<sub>3</sub>, the reduction reaction is inhibited by oxygen, favoring a necrotic response in the centrilobular area [7].

Liver cell necrosis can evoke a range of responses within the liver that depend on the severity of insult. These responses range from regeneration of necrotic tissue with restoration of full liver function to concomitant loss of liver function, liver failure and death [8]. In the process, necrosis may trigger the development of other liver diseases. Recurring bouts of necrosis and repair may result in disruption of the structure of the liver and result in subacute hepatitis, chronic hepatitis or even cirrhosis [9]. In this process, inflammatory cells stimulate the deposition of collagen around hepatocytes causing alteration in hepatic function and blood flow [10]. There is also evidence that necrosis may play a role in the induction of early hepatocellular carcinoma through compensatory liver regeneration. Diethylnitrosamine and Fumonisin B(1) are two examples of compounds that show evidence of producing hepatocellular carcinoma in rats through a sequence of events that begin with necrosis [11-13]. Thus the benefit of obtaining

marker genes predictive of hepatic zone 3 necrosis stem from its participation in the pathogenesis of other liver diseases as well as it being an early indicator of hepatic toxicity.

#### SUMMARY OF THE INVENTION

The invention is based in part on the discovery that certain nucleic acids are differentially  
5 expressed in liver cells or liver tissue of animals treated with toxic compounds. These  
differentially expressed nucleic acids include novel sequences and nucleic acids sequences that,  
while previously described, have not heretofore been identified as associated with toxicity and  
are collectively referred to herein as "TOXMARKER nucleic acids" or "TOXMARKER  
polynucleotides" and the corresponding encoded polypeptides are referred to as "TOXMARKER  
10 polypeptides" or "TOXMARKER proteins". The TOXMARKER genes are useful in high  
throughput screening of potential therapeutic compounds for toxicity.

In one aspect the invention provides methods of predicting the hepatotoxicity of a test  
agent. Hepatotoxicity is predicted by determining the level of expression of a toxicity-associated  
gene in a cell exposed to a test agent. The level of expression of the toxicity-associated gene is  
15 compared to the level of expression of the toxicity-associated gene in a control population  
exposed to a control agent. A test agent is predicted to be toxic if an alteration (e.g., increase or  
decrease) in the level of expression in the cell exposed to the test agent compared to the control  
population is identified.

Also provided by the invention are methods of screening a test agent for inducing changes  
20 in gene expression associated with a toxic agent. An agent is screened for inducing changes in  
gene expression associated with a toxic agent by determining the level of expression of a  
toxicity-associated gene in a cell exposed to a test agent. The level of expression of the toxicity-  
associated gene is compared to the level of expression of the toxicity-associated gene in a control  
population exposed to a control agent.

25 The alteration is statistically significant. By statistically significant is meant that the  
alteration is greater than what might be expected to happen by chance alone. Statistical  
significance is determined by method known in the art. An alteration is statistically significant if  
the p-value is at least 0.05. Preferably, the p-value is 0.04, 0.03, 0.02, 0.01, 0.005, 0.001 or less.

By toxicity- associated gene is meant a gene that is characterized by a level of expression which differs in a cell exposed to a toxic compound compared to a control population. A toxicity-associated gene includes for example TOXMARKER 1-131. Preferably, the toxicity-associated gene is the genes listed on Table 5. More preferably, the toxicity-associated gene is  
5 TOXMARKER 42, 59, 65, 66, 71, 76, and 97

A control population is a for example a cell not exposed to a toxic agent. Optionally, the control population is exposed to a control agent. A control agent is an agent that does not elicit a histology associated with a condition associated with liver toxicity such as Cholestasis; Steatosis; Reactive Inflammation; Necrosis, *e.g.*, zone 3, general or multifocal; Genotoxic Carcinogenesis;  
10 Hepatocellular Hypertrophy; Non-Genotoxic Carcinogenesis; Apoptosis and Kupffer Cell Aggregation. Exemplary control agents are those listed in Table 1 and Table 2 below. A control level is a single expression pattern derived from a single control population or from a plurality of expression patterns. For example, the control level can be a database of expression patterns from previously tested cells.

15 The test cell is provided *in vitro*. Alternatively, the test cell is provided *ex vivo* or *in vivo* from a mammalian subject. The test cell is derived from liver tissue, such as for example a hepatocyte. Alternatively, the test cell is a subject derived cell sample. The subject derived tissue sample is any tissue from a test subject.

Expression is determined by for example detecting hybridization, *e.g.*, on a chip, of a  
20 toxicity-associated gene probe to a gene transcript of the test cell.

The invention also provides a zone 3 necrosis reference expression profile of a gene expression level two or more of TOXMARKER 1-132. For example, the reference profile contains the expression levels of TOXMARKER 1-132. Alternatively, the reference profile contains the expression levels of TOXMARKER genes listed on Table 5. Preferably, the  
25 reference profile contains the expression levels of TOXMARKER 42, 59, 65, 66, 71, 76, and 97

The invention also provides a kit with a detection reagent which binds to two or more TOXMARKER nucleic acid sequences or which binds to a gene product encoded by the nucleic acid sequences. Also provided is an array of nucleic acids, *e.g.* oligonucleotides that binds to two or more TOXMARKER nucleic acids. For example, the array contains oligonucleotides that



bind TOXMARKER 1-132. Alternatively, the array contains oligonucleotides that bind the TOXMARKER genes listed on Table 5. Preferably, the array contains oligonucleotides that bind TOXMARKER 42, 59, 65, 66, 71, 76, and 97. Most preferably, the array contains oligonucleotides that binds at least five TOXMARKER genes listed on Table 5, where the  
5 collection of TOXMARKER genes predict toxicity to a confidence level of a p-value of at least 0.05 or less.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be  
10 used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

15 Other features and advantages of the invention will be apparent from the following detailed description and claims.

#### **BRIEF DESCRIPTION OF THE DRAWING**

FIG. 1 is a schematic of the hepatotoxicity prediction screening method of the invention.

FIG. 2 is an illustration of a chart showing the linear discriminant model.

#### **20 DETAILED DESCRIPTION OF THE INVENTION**

The present invention is based in part on the discovery of changes in expression patterns of multiple nucleic acid sequences in response to compounds known to elicit a histopathologic condition (*i.e.*, a pathologic change to the liver visible on examination that causes the liver to function less effectively) associated with liver toxicity. The compounds are for example,  
25 compounds listed in Table 1 or 2 below. Examples of histopathologic conditions include, Cholestasis; Steatosis; Reactive Inflammation; Necrosis, *e.g.*, zone 3, general or multifocal; Genotoxic Carcinogenesis; Hepatocellular Hypertrophy; Non-Genotoxic Carcinogenesis;

Apoptosis and Kupffer Cell Aggregation. The histopathological conditions are identified by methods known in the art. For example, zone 3 necrosis is identified by cholestasis and hypertrophy. The identification of genes that are differentially expressed in response to toxic compounds are useful in screening potential therapeutic compositions for toxicity.

5           The genes whose expression levels are modulated (*i.e.*, increased or decreased) in response to exposure to a toxic compound are summarized in Tables 3-5 (see EXAMPLES 4 and 5) and are collectively referred to herein as "toxicity-associated gene", "TOXMARKER nucleic acids" or "TOXMARKER polynucleotides" and the corresponding encoded polypeptides are referred to as "TOXMARKER polypeptides" or "TOXMARKER proteins." Unless indicated  
10 otherwise, "TOXMARKER" or "toxicity-associated gene" is meant to refer to any of the sequences disclosed herein.

          For a given TOXMARKER sequence, its expression can be measured in the methods described herein. For previously described sequences, database accession numbers are provided. This information allows for one of ordinary skill in the art to deduce information necessary for  
15 detecting and measuring expression of the TOXMARKER nucleic acid sequences.

## GENERAL METHODS

          The TOXMARKER nucleic acids and encoded polypeptides can be identified using the information provided in the EXAMPLES below. In some embodiments, the TOXMARKER nucleic acids and polypeptides correspond to the nucleic acids or polypeptides which include the  
20 various sequences (referenced by SEQ ID NOs) disclosed for each TOXMARKER.

          The invention includes providing a test cell population which includes at least one cell that is capable of expressing one or more of the sequences TOXMARKER 1-132. By "capable of expressing" is meant that the gene is present in an intact form in the cell and can be expressed. Expression of one, some, or all of the TOXMARKER sequences is then detected, if present, and,  
25 preferably, measured to yield an expression profile, *e.g.*, subject expression profile or a test cell expression profile. By "expression profile" is meant a pattern of the level of expression of at least two toxicity-associated genes.

          "Similarity of expression profile" is a similarity of expression profile between two

samples exists when the linear combination of the genes in the profile has a linear discriminant score that is more similar to one of the training classes than the other. Linear discriminant analysis (LDA) identifies a linear combination of markers that best separates the defined classes. In the training data (*i.e.*, control population) of this invention linear discriminant score could be  
5 determined by the following equation:

$$\text{Linear disc. score} = a\text{Gene1} + b\text{Gene2} + \dots + n\text{GeneN}$$

where a, b ... n are the coefficients identified by least squares that best separate the phenotypes under investigation. Thus, similarity in expression profile is a similarity in gene combinations. Interpretation of raw data is difficult since the samples are plotted in more than 3  
10 dimensions, one dimension for each gene, which makes it difficult to visualize the data. LDA compresses this information into a single dimension.

By “toxicity-associated gene” is meant a gene, which the level of expression differs in a cell or subject exposed to a known toxic compound as compared to a cell or subject not exposed to a toxic compound (*i.e.*, control). Preferably, the TOXMARKER genes 42 (IFNAR-2), 59  
15 (Transaldolase), 65 (Clp-1), 66 (Hex), 71 (cszr\_204152648\_191521095), 76 (scr\_gb-aa899865\_3), and 97 (scr\_gb-bm986259\_1).

Using sequence information provided by the database entries for the known sequences, or the sequence information provided herein for the newly described sequences, expression of the TOXMARKER sequences are detected (if present) and measured using techniques well known to  
20 one of ordinary skill in the art. For example, sequences within the sequence database entries corresponding to TOXMARKER sequences, or within the sequences disclosed herein, can be used to construct probes for detecting TOXMARKER RNA sequences in, *e.g.*, northern blot hybridization analyses or methods which specifically, and, preferably, quantitatively amplify specific nucleic acid sequences. As another example, the sequences can be used to construct  
25 primers for specifically amplifying the TOXMARKER sequences in, *e.g.*, amplification-based detection methods such as reverse-transcription based polymerase chain reaction. When alterations in gene expression are associated with gene amplification or deletion, sequence comparisons in test and reference populations can be made by comparing relative amounts of the examined DNA sequences in the test and reference cell populations.

Expression of the genes disclosed herein can be measured at the RNA level using any method known in the art. For example, northern hybridization analysis using probes which specifically recognize one or more of these sequences can be used to determine gene expression. Alternatively, expression can be measured using reverse-transcription-based PCR assays, *e.g.*,  
5 using primers specific for the differentially expressed sequences.

Expression is also measured at the protein level, *i.e.*, by measuring the levels of polypeptides encoded by the gene products described herein. Such methods are well known in the art and include, *e.g.*, immunoassays based on antibodies to proteins encoded by the genes.

Expression levels of one or more of the TOXMARKER sequences in the test cell  
10 population are then compared to expression levels of the sequences in one or more cells from a reference (*i.e.*, control) cell population. If desired, a reference expression profile is generated. A reference profile is a single expression pattern derived from a single reference population or from a plurality of expression patterns. For example, the reference cell population can be a database of expression patterns from previously tested cells for which one of the herein-described parameters  
15 or conditions (*e.g.*, toxicity) is known.

The reference profile is obtained from the training data. Training data is a collection of data from the *in vitro* or *in vivo* samples that were exposed to compounds that produce a known pathology. (*i.e.*, pathology present or pathology absent) Profile is defined here to indicate the absolute estimate of the expression level of any one TOXMARKER gene fragment (*e.g.*  
20 Intensity).

Expression of sequences in test and reference populations of cells are compared using any art-recognized method for comparing expression of nucleic acid sequences. For example, expression can be compared using GENECalling<sup>®</sup> methods as described in US Patent No. 5,871,697 and in Shimkets et al., *Nat. Biotechnol.* 17:798-803.

25 In various embodiments, the expression of one or more sequences encoding genes of related function, as listed in Tables 3-5, is compared. In various embodiments, the expression of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 35, 40, 50, 100 or all of the sequences represented by TOXMARKER 1-132 are measured. If desired, expression of these sequences can be measured

along with other sequences whose expression is known to be altered according to one of the herein described parameters or conditions.

The reference cell population includes one or more cells for which the compared parameter is known. The compared parameter can be, *e.g.*, toxic agent expression status. By “toxic agent expression status” is meant that it is known whether the reference cell has had contact with a toxic agent. Whether or not comparison of the gene expression profile in the test cell population to the reference cell population reveals the presence, or degree, of the measured parameter depends on the composition of the reference cell population. For example, if the reference cell population is composed of cells that have not been treated with a known toxic agent, a similar gene expression level in the test cell population and a reference cell population indicates the test agent is not a toxic agent. Conversely, if the reference cell population is made up of cells that have been treated with a toxic agent, a similar gene expression profile between the test cell population and the reference cell population indicates the test agent is a toxic agent.

In various embodiments, a TOXMARKER sequence in a test cell population is considered comparable in expression level to the expression level of the TOXMARKER sequence if its expression level varies within a factor of 2.0, 1.5, or 1.0 fold to the level of the TOXMARKER transcript in the reference cell population. In various embodiments, a TOXMARKER sequence in a test cell population can be considered altered in levels of expression if its expression level varies from the reference cell population by more than 3.0, 4.0, 5.0 or more fold from the expression level of the corresponding TOXMARKER sequence in the reference cell population.

If desired, comparison of differentially expressed sequences between a test cell population and a reference cell population can be done with respect to a control nucleic acid whose expression is independent of the parameter or condition being measured. Expression levels of the control nucleic acid in the test and reference nucleic acid can be used to normalize signal levels in the compared populations.

In some embodiments, the test cell population is compared to multiple reference cell populations. Each of the multiple reference populations may differ in the known parameter.

Thus, a test cell population may be compared to a first reference cell population known to have been exposed to a toxic agent, as well as a second reference population known to have not been exposed to a toxic agent.

5       The test cell population that is exposed to, *i.e.*, contacted with, the test toxic agent can be any number of cells, *i.e.*, one or more cells, and can be provided *in vitro*, *in vivo*, or *ex vivo*. The test cell is obtained from a bodily fluid, *e.g.*, biological fluid (such as blood, serum, urine, saliva, milk, ductal fluid, or tears). For example, the test cell is purified from blood or another tissue, *i.e.*, liver tissue.

10       In other embodiments, the test cell population can be divided into two or more subpopulations. The subpopulations can be created by dividing the first population of cells to create as identical a subpopulation as possible. This will be suitable, in, for example, *in vitro* or *ex vivo* screening methods. In some embodiments, various subpopulations can be exposed to a control agent, and/or a test agent, multiple test agents, or, *e.g.*, varying dosages of one or multiple test agents administered together, or in various combinations.

15       Preferably, cells in the reference cell population are derived from a tissue type as similar as possible to the test cell, *e.g.*, liver tissue. Alternatively the cells are derived from an established cell line. Preferably, the cell is a hepatocyte. In some embodiments, the control cell is derived from the same subject as the test cell, *e.g.*, from a region proximal to the region of origin of the test cell. In other embodiments, the reference cell population is derived from a  
20       plurality of cells. For example, the reference cell population can be a database of expression patterns from previously tested cells for which one of the herein-described parameters or conditions (toxic agent expression status) is known.

      The test agent can be a compound not previously described or can be a previously known compound but which is not known to be a toxic agent.

25       The subject is preferably a mammal. The mammal can be, *e.g.*, a human, non-human primate, mouse, rat, dog, cat, horse, or cow.

## PREDICTION OF TOXICITY

In one aspect, the invention provides a method of predicting the toxicity e.g., hepatotoxicity of a test agent or identifying a toxic agents, e.g., a hepatotoxic agent. The method is an *in vivo* method. Alternatively, the method is an *in vitro* method.

5 Differences in the genetic makeup of individuals can result in differences in their relative abilities to metabolize various drugs. Accordingly, the differentially expressed TOXMARKER sequences disclosed herein allow for a putative therapeutic or prophylactic agent to be tested in a test cell population from a selected subject in order to predict if the agent causes toxicity in the subject.

10 By predicting the toxicity is meant that the test compound is more likely to be hepatotoxic than not be hepatotoxic. Hepatotoxicity is predicted by determining the level of expression of a toxicity-associated gene in a cell exposed to a test agent. The level of expression of the toxicity-associated gene is compared to the level of expression of the toxicity-associated gene in a control population exposed to a control agent. A test agent is predicted to be toxic if an alteration (e.g.,  
15 increase or decrease) in the level of expression in the cell exposed to the test agent compared to the control population is identified.

The toxicity-associated gene is for example TOXMARKER 1-132. Alternatively, the toxicity-associated gene is the TOXMARKER genes listed on Table 5. Optionally, the toxicity-associated gene is TOXMARKER 42, 59, 65, 66, 71, 76, and 97. The toxicity-associated gene is  
20 a nucleic acid sequences homologous to those listed in Tables 3-5 as TOXMARKER 1-132. The sequences need not be identical to sequences including TOXMARKER 1-132, as long as the sequence is sufficiently similar that specific hybridization can be detected. Preferably, the cell includes sequences that are identical, or nearly identical to those identifying the TOXMARKER nucleic acids shown in Tables 3-5.

25 By hepatotoxicity is meant that the compound causes a histopathological change in the live tissue associated with toxicity. By "toxicity" is meant that the agent is damaging or destructive to liver when administered to a subject. Damage to the liver is measured for

example, histologically. Hepatotoxicity is determined, for example as described in the examples below.

The cell population is contacted *in vitro*, or *in vivo*. Optionally, the cell population is contacted *ex vivo* with the agent or activated form of the agent.

5            Expression of the nucleic acid sequences in the test cell population is then compared to the expression of the nucleic acid sequences in a control population, which is a cell population that has not been exposed to the test agent, or, in some embodiments, a cell population exposed to the test agent. Comparison can be performed on test and reference samples measured concurrently or at temporally distinct times. An example of the latter is the use of compiled  
10   expression information, *e.g.*, a sequence database, which assembles information about expression levels of known sequences following administration of various agents. For example, alteration of expression levels following administration of test agent can be compared to the expression changes observed in the nucleic acid sequences following administration of a control agent. A control agent is a compound that elicits the histopathology. Alternatively, the control agent is a  
15   compound that does not elicit the histopathology. Exemplary control compounds are listed in Tables 1 and 2.

An alteration in expression of the nucleic acid sequence in the test cell population compared to the expression of the nucleic acid sequence in the control cell population that has not been exposed to the test agent indicates the test agent is a toxic agent.

20            The alteration is statistically significant. By statistically significant is meant that the alteration is greater than what might be expected to happen by chance alone. Statistical significance is determined by method known in the art. For example statistical significance is determined by p-value. The p-value is a measure of probability that a difference between groups during an experiment happened by chance. ( $P(z \geq z_{\text{observed}})$ ). For example, a p-value of 0.01 means  
25   that there is a 1 in 100 chance the result occurred by chance. The lower the p-value, the more likely it is that the difference between groups was caused by treatment. An alteration is statistically significant if the p-value is at least 0.05. Preferably, the p-value is 0.04, 0.03, 0.02, 0.01, 0.005, 0.001 or less.



The invention also includes a toxic agent identified according to this screening method.

The differentially expressed TOXMARKER sequences identified herein also allow for the toxicity of a toxic agent to be determined or monitored. In this method, a test cell population from a subject is exposed to a test agent, *i.e.* a. toxic agent. If desired, test cell populations can  
5 be taken from the subject at various time points before, during, or after exposure to the test agent. Expression of one or more of the TOXMARKER sequences, *e.g.*, TOXMARKER: 1-132, in the cell population is then measured and compared to a control population which includes cells whose toxic agent expression status is known.

### KITS

10 The invention also includes a TOXMARKER-detection reagent, *e.g.*, nucleic acids that specifically identify one or more TOXMARKER nucleic acids by having homologous nucleic acid sequences, such as oligonucleotide sequences, complementary to a portion of the TOXMARKER nucleic acids or antibodies to proteins encoded by the TOXMARKER nucleic acids packaged together in the form of a kit. The oligonucleotides are fragments of the the  
15 TOXMARKER genes. For example the oligonucleitides are 200, 150, 100, 50, 25, 10 or less nucleotides in length. The kit may contain in separate containers a nucleic acid or antibody (either already bound to a solid matrix or packaged separately with reagents for binding them to the matrix) , control formulations (positive and/or negative), and/or a detectable label. Instructions (*e.g.*, written, tape, VCR, CD-ROM, etc.) for carrying out the assay may be included  
20 in the kit. The assay may for example be in the form of a Northern hybridization or a sandwich ELISA as known in the art.

For example, TOXMARKER detection reagent, is immobilized on a solid matrix such as a porous strip to form at least one TOXMARKER detection site. The measurement or detection region of the porous strip may include a plurality of sites containing a nucleic acid. A test strip  
25 may also contain sites for negative and/or positive controls. Alternatively, control sites are located on a separate strip from the test strip. Optionally, the different detection sites may contain different amounts of immobilized nucleic acids, *i.e.*, a higher amount in the first detection site and lesser amounts in subsequent sites. Upon the addition of test sample, the

number of sites displaying a detectable signal provides a quantitative indication of the amount of TOXMARKER present in the sample. The detection sites may be configured in any suitably detectable shape and are typically in the shape of a bar or dot spanning the width of a test strip.

Alternatively, the kit contains a nucleic acid substrate array comprising one or more nucleic acid sequences. The nucleic acids on the array specifically identify one or more nucleic acid sequences represented by TOXMARKER 1-132. In various embodiments, the expression of 2, 3,4, 5, 6, 7,8, 9, 10, 15, 20, 25, 40 or 50 or more of the sequences represented by TOXMARKER 1-132. are identified by virtue of binding to the array. The substrate array can be on, *e.g.*, a solid substrate, *e.g.*, a “chip” as described in U.S. Patent No.5,744,305.

## 10    **ARRAYS AND PLURALITIES**

The invention also includes a nucleic acid substrate array comprising one or more nucleic acid sequences. The nucleic acids on the array specifically identify one or more nucleic acid sequences represented by TOXMARKER 1-132. In various embodiments, the expression of 2, 3,4, 5, 6, 7,8, 9, 10, 15, 20, 25, 40 or 50 or more of the sequences represented by TOXMARKER 1-132 are identified.

The nucleic acids in the array can identify the enumerated nucleic acids by, *e.g.*, having homologous nucleic acid sequences, such as oligonucleotide sequences, complementary to a portion of the recited nucleic acids. The substrate array can be on, *e.g.*, a solid substrate, *e.g.*, a “chip” as described in U.S. Patent No.5,744,305.

The invention also includes an isolated plurality (*i.e.*, a mixture of two or more nucleic acids) of nucleic acid sequences. The nucleic acid sequence can be in a liquid phase or a solid phase, *e.g.*, immobilized on a solid support such as a nitrocellulose membrane. The plurality typically includes one or more of the nucleic acid sequences represented by TOXMARKER 1-132. In various embodiments, the plurality includes 2, 3,4, 5, 6, 7,8, 9, 10, 15, 20, 25, 40 or 50 or more of the sequences represented by TOXMARKER 1-132.

## Nucleic Acids

One aspect of the invention pertains to isolated nucleic acid molecules that encode TOXMARKER proteins or biologically active portions thereof, as well as nucleic acid fragments sufficient for use as hybridization probes to identify TOXMARKER-encoding nucleic acids (*e.g.*, TOXMARKER mRNA) and fragments for use as PCR primers for the amplification or mutation of TOXMARKER nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

"Probes" refer to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as about, *e.g.*, 6,000 nt, depending on use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are usually obtained from a natural or recombinant source, are highly specific and much slower to hybridize than oligomers. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

An "isolated" nucleic acid molecule is one that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated TOXMARKER nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived (*e.g.*, testis, lung, B-cells). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1-171 or a complement of any of these nucleotide sequences, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:  
5 1-171 as a hybridization probe, TOXMARKER molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2<sup>nd</sup> Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

10 A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to TOXMARKER nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*,  
15 using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or  
20 complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO: 1-171, or a complement thereof. Oligonucleotides may be chemically synthesized and may be used  
25 as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO: 1-171. In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO: 1-

171, or a portion of this nucleotide sequence. A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NO: 1-171 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO: 1-171 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown in SEQ ID NO: 1-171, thereby forming a stable duplex.

As used herein, the term “complementary” refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term “binding” means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, Von der Waals, hydrophobic interactions, etc. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Moreover, the nucleic acid molecule of the invention can comprise only a portion of the nucleic acid sequence of SEQ ID NO: 1-171, *e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically active portion of TOXMARKER.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 30%, 50%, 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A “homologous nucleic acid sequence” or “homologous amino acid sequence,” or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of TOXMARKER polypeptide. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the present invention, homologous nucleotide sequences include nucleotide sequences encoding for a TOXMARKER polypeptide of species other than humans, including, but not limited to, mammals, and thus can include, *e.g.*, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the nucleotide sequence encoding human TOXMARKER protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO: 1-171, as well as a polypeptide having TOXMARKER activity. Biological activities of the TOXMARKER proteins are described below. A homologous amino acid sequence does not encode the amino acid sequence of a human TOXMARKER polypeptide.

An TOXMARKER polypeptide is encoded by the open reading frame ("ORF") of a TOXMARKER nucleic acid. An "open reading frame" ("ORF") corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding  
5 sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a bona fide cellular protein, a minimum size requirement is often set, for example, a stretch of DNA that would encode a protein of 50 amino  
10 acids or more.

The nucleotide sequence determined from the cloning of the human TOXMARKER gene allows for the generation of probes and primers designed for use in identifying and/or cloning TOXMARKER homologues in other cell types, *e.g.* from other tissues, as well as TOXMARKER homologues from other mammals. The probe/primer typically comprises  
15 substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO: 1-171, or an anti-sense strand nucleotide sequence of SEQ ID NO: 1-171 or of a naturally occurring mutant of SEQ ID NO: 1-171.

20 Probes based on the human TOXMARKER nucleotide sequence can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissue which misexpress a  
25 TOXMARKER protein, such as by measuring a level of a TOXMARKER-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting TOXMARKER mRNA levels or determining whether a genomic TOXMARKER gene has been mutated or deleted.

"A polypeptide having a biologically active portion of TOXMARKER" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a

polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically active portion of TOXMARKER" can be prepared by isolating a portion of SEQ ID NO: 1-171 that encodes a polypeptide having a TOXMARKER biological activity (the biological activities of the TOXMARKER proteins are described below), expressing the encoded portion of TOXMARKER protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of TOXMARKER.

### **TOXMARKER variants**

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO: 1-171 due to degeneracy of the genetic code and thus encode the same TOXMARKER protein as that encoded by the nucleotide sequence shown in SEQ ID NO: 1-171.

In addition to the human TOXMARKER nucleotide sequence shown in SEQ ID NO: 1-171 it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of TOXMARKER may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the TOXMARKER gene may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a TOXMARKER protein, preferably a mammalian TOXMARKER protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the TOXMARKER gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in TOXMARKER that are the result of natural allelic variation and that do not alter the functional activity of TOXMARKER are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding TOXMARKER proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO: 1-171 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the TOXMARKER cDNAs of the invention can be isolated based on their homology to the human TOXMARKER nucleic acids disclosed herein



using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a soluble human TOXMARKER cDNA can be isolated based on its homology to human membrane-bound TOXMARKER. Likewise, a membrane-bound human TOXMARKER cDNA can be isolated  
5 based on its homology to soluble human TOXMARKER.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-171. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000 or 1250 nucleotides in length. In  
10 another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding TOXMARKER proteins derived from species  
15 other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other  
20 sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of  
25 the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at  $T_m$ , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides

(*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 5 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% 10 Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO: 1-171 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a 15 nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-171, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's 20 solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known in the art. See, *e.g.*, Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

25 In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-171, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured

salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR

TRA 1832808v1

- 5 BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981, *Proc Natl Acad Sci USA* 78: 6789-6792.

## EXAMPLES

### Example 1: Induction of Zone 3 Necrosis *in vivo*

- 10 Over 100 compounds were chosen based on evidence that they elicit one of eleven selected histopathology subtypes. The criteria for inclusion of a compound into the nongenotoxic carcinogens mode included evidence of parenchymal changes and an increase in mitosis *in vivo*. Compounds assigned to this group must also have strong historical documentation. Compounds from other pathology modes were not added to this histopathology subtype. The compounds
- 15 included in nongenotoxic carcinogenesis can be seen in Table 1. Each compound was delivered orally on a daily basis at a high dose (tox dose) and a 1/10 low dose (mode dose) for up to 14 days. Five male rats/dose/time were randomly assigned to sacrifice on days 1, 3, 7, and 14. In order to best identify genes characteristic of the histopathology subtype, total RNA for all rat livers from a given dose time point were pooled and converted to mRNA and cDNA for
- 20 GeneCalling®. In GeneCalling, the cDNA is cut with a battery of restriction enzyme pairs in different combinations followed by amplification by PCR using specific primers linked to specific adaptors. After gel electrophoresis, the resulting fragments are identified based on the inherent information in the cDNA fragment: The flanking restriction site sequences on the ends, the size of the fragment and the species (and sometimes the tissue) origin of the DNA. This
- 25 information is used to query public and proprietary databases. The fragments that do not match any sequences in the database are isolated, sequenced and identified as novel.

**Table 1: Zone 3 Necrosis *in vivo***

Compound	Vehicle	Dose	Concentration	Time Points	Pathology Present	Pathology Absent
----------	---------	------	---------------	-------------	-------------------	------------------

1,3-Dibromobenzene	Corn Oil	High	600 mg/kg/d	1,3,7,14 d	1 d	
1,3-Dibromobenzene	Corn Oil	Low	60 mg/kg/d	1,3,7,14 d		
1,4-dichlorobenzene	Corn Oil	High	300 mg/kg/d	1,3,7,14 d		
1,4-dichlorobenzene	Corn Oil	Low	30 mg/kg/d	1,3,7,14 d		1,3,7,14 d
17 $\alpha$ -Ethinyl-19-nortestosterone	Corn Oil	High	30 mg/kg/d	1,3,7,14 d		
17 $\alpha$ -Ethinyl-19-nortestosterone	Corn Oil	Low	3 mg/kg/d	1,3,7,14 d		
2,4-diaminotoluene	Methylcellulose	High	15 mg/kg/d	1,3,7,14 d		
2,4-diaminotoluene	Methylcellulose	Low	1.5 mg/kg/d	1,3,7,14 d		
2-acetylaminofluorene	Methylcellulose	High	12 mg/kg/d	1,3,7,14 d		1,3,7,14 d
2-acetylaminofluorene	Methylcellulose	Low	1.2 mg/kg/d	1,3,7,14 d		
2-acetylaminofluorene	Methylcellulose	High	120 mg/kg (ip)	6,12,24,48 h		
2-acetylaminofluorene	Methylcellulose	Low	12 mg/kg (ip)	6,12,24,48 h		
2-nitrofluorene	Corn oil	High	44 mg/kg/d	1,3,7,14 d		
2-nitrofluorene	Corn oil	Low	4.4 mg/kg/d	1,3,7,14 d		
3-methyl-4-(dimethylamino)azobenzene	Methylcellulose	High	36 mg/kg/d	1,3,7,14 d		
3-methyl-4-(dimethylamino)azobenzene	Methylcellulose	Low	3.6 mg/kg/d	1,3,7,14 d		1,3,7,14 d
3-methylcholanthrene	Corn Oil	High	25 mg/kg/d (ip)	1,3,7,14 d		
3-methylcholanthrene	Corn Oil	Low	2.5 mg/kg/d (ip)	1,3,7,14 d		1,3,7,14 d
Acetamide	Methylcellulose	High	3000 mg/kg/d	1,3,7,14 d		
Acetamide	Methylcellulose	Low	300mg/kg/d	1,3,7,14 d		
Acetaminophen	Saline	High	4.25 g/kg (po)	6,12,24,48 h	24,48 h	
Acetaminophen	Saline	Low	425 mg/kg (po)	6,12,24,48 h		
Aflatoxin B1	Methylcellulose	High	0.24 mg/kg/d	1,3,7,14 d		
Aflatoxin B1	Methylcellulose	Low	0.024 mg/kg/d	1,3,7,14 d		
Allyl Alcohol	Methylcellulose	High	36 mg/kg/d	1,3,7,14 d		
Allyl Alcohol	Methylcellulose	Low	3.6 mg/kg/d	1,3,7,14 d		
Allyl Formate	Corn oil	High	94.8 mg/kg (ip)	3, 6, 12, 24 h		
Allyl Formate	Corn oil	Low	9.48 mg/kg (ip)	3, 6, 12, 24 h		
Amiodarone	Methylcellulose	High	500 mg/kg/d	1,3,7,14 d		
Amiodarone	Methylcellulose	Low	50 mg/kg/d	1,3,7,14 d		
ANIT	Corn Oil	High	60 mg/kg/d	1,3,7,14 d		
ANIT	Corn Oil	Low	6 mg/kg/d	1,3,7,14 d		

Azaserine	Saline	High	100 mg/kg (ip)	1,3,7,14 d	1,3 d	
Azaserine	Saline	Low	10 mg/kg (ip)	1,3,7,14 d		
BCNU	Corn Oil	High	20 mg/kg/d	1,3,7,14 d		
BCNU	Corn Oil	Low	2 mg/kg/d	1,3,7,14 d		
BHT	Corn Oil	High	500 mg/kg/d	1,3,7,14 d		
BHT	Corn Oil	Low	50 mg/kg/d	1,3,7,14 d		
Bromobenzene	Saline	High	1200 mg/kg (ip)	6,12,24,48 h	12,48 h	
Bromobenzene	Saline	Low	120 mg/kg (ip)	6,12,24,48 h		
C.I. Direct Black	Corn oil	High	146 mg/kg/d	1,3,7,14 d		
C.I. Direct Black	Corn oil	Low	14.6 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Cadmium chloride	Saline	High	3.9 mg/kg (iv)	6,12,24,48 h		
Cadmium chloride	Saline	Low	0.39 mg/kg (iv)	6,12,24,48 h		
Carbamazepine	Methylcellulose	High	200 mg/kg/d	1,3,7,14 d		
Carbamazepine	Methylcellulose	Low	20 mg/kg/d	1,3,7,14 d		
CCl <sub>4</sub>	Corn Oil	High	50 mg/kg/d	1,3,7,14 d	7 d	
CCl <sub>4</sub>	Methylcellulose	High	956 mg/kg (ip)	6,12,24,48 h	6,12,48 h	
CCl <sub>4</sub>	Corn Oil	Low	5 mg/kg/d	1,3,7,14 d		
CCl <sub>4</sub>	Methylcellulose	Low	95.6 mg/kg (ip)	6,12,24,48 h		
CCNU	Corn Oil	High	20 mg/kg/d	1,3,7,14 d		
CCNU	Corn Oil	Low	2 mg/kg/d	1,3,7,14 d		
Cefuroxime	Methylcellulose	Safe	125 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Cerium (III) Chloride	Saline	High	10 mg/kg (iv)	1,2,3,4 d	1,2,3,4 d	
Cerium (III) Chloride	Saline	Low	1 mg/kg (iv)	1,2,3,4 d		
Chlordane	Corn Oil	High	25 mg/kg/d	1,3,7,14 d		
Chlordane	Corn Oil	Low	2.5 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Ciprofibrate	Corn Oil	High	10 mg/kg/d	1,3,7,14 d		
Ciprofibrate	Corn Oil	Low	1 mg/kg/d	1,3,7,14 d		
Ciprofloxacin	Methylcellulose	Safe	40 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Clofibrate	Methylcellulose	High	300 mg/kg/d	1,3,7,14 d		
Clofibrate	Methylcellulose	Low	30 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Clonidine	Methylcellulose	Safe	0.1 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Colchicine	Saline	High	5 mg/kg x 2 (ip)	6,12,24,48 h		
Colchicine	Saline	Low	0.5 mg/kg x 2 (ip)	6,12,24,48 h		
Concanavalin A	Saline	High	20 mg/kg (iv)	6,12,24,48 h		

Concanavalin A	Saline	Low	2 mg/kg (iv)	6,12,24,48 h		
Corn Oil		Control		1,3,7,14 d		1,3,7,14 d
Corn Oil (ip)		Control		1,3,7,14 d		1,3,7,14 d
Corn Oil (ip)		Control		3,6,12,24 h		3,6,12,24 h
Coumarin	Corn Oil	High	150 mg/kg/d	1,3,7,14 d	3 d	
Coumarin	Corn Oil	Low	15 mg/kg/d	1,3,7,14 d		
CTFT	Corn Oil	High	1 g/kg/d	1,3,7,14 d		
CTFT	Corn Oil	Low	100 mg/kg/d	1,3,7,14 d		
Cyclosporine A	Corn Oil	High	50 mg/kg/day	1,3,7,14 d		
Cyclosporine A	Corn Oil	Low	5 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Cyproterone acetate	Corn Oil	High	100 mg/kg/d	1,3,7,14 d		
Cyproterone acetate	Corn Oil	Low	10 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Dehydroepiandrosterone	Methylcellulose	High	600 mg/kg/d	1,3,7,14 d		
Dehydroepiandrosterone	Methylcellulose	Low	60 mg/kg/d	1,3,7,14 d		
Deoxycholic Acid	Methylcellulose	High	300 mg/kg/d	1,3,7,14 d		
Deoxycholic Acid	Methylcellulose	Low	30 mg/kg/d	1,3,7,14 d		
Dexamethasone	Corn oil	High	50 mg/kg/d (ip)	1,3,7,14 d		
Dexamethasone	Corn oil	Low	5 mg/kg/d (ip)	1,3,7,14 d		1,3,7,14 d
D-galactosamine	Saline	High	500 mg/kg (ip)	6,12,24,48 h		
D-galactosamine	Saline	Low	50 mg/kg (ip)	6,12,24,48 h		
Di(2-Ethylhexyl) phthalate	Methylcellulose	High	1200 mg/kg/d	1,3,7,14 d		
Di(2-Ethylhexyl) phthalate	Methylcellulose	Low	120 mg/kg/d	1,3,7,14 d		
Dibutyltin chloride	Corn oil	High	20 mg/kg/d	1,3,7,14 d		
Dibutyltin chloride	Corn oil	Low	2 mg/kg/d	1,3,7,14 d		
Dichloropropane	Corn oil	High	1000 mg/kg/d	1,3,7,14 d	1,3 d	
Dichloropropane	Corn oil	Low	100 mg/kg/d	1,3,7,14 d		
Diethylnitrosamine	Saline	High	150 mg/kg (ip)	1,3,7,14 d	1,3,7 d	
Diethylnitrosamine	Saline	Low	15 mg/kg (ip)	1,3,7,14 d		
Diethylstilbestrol	Methylcellulose	High	10 mg/kg/d	1,3,7,14 d		
Diethylstilbestrol	Methylcellulose	Low	1 mg/kg/d	1,3,7,14 d		
Dimethylformamide	Saline	High	850 mg/kg (ip)	6,12,24,48 h	48 h	
Dimethylformamide	Saline	Low	85 mg/kg (ip)	6,12,24,48 h		
Dimethylnitrosamine	Corn Oil	High	4 mg/kg/d	1,3,7,14 d	7,14 d	
Dimethylnitrosamine	Corn Oil	Low	0.4 mg/kg/d	1,3,7,14 d		

Diquat	Saline	High	36 mg/kg (ip)	6,12,24,48 h		
Diquat	Saline	Low	3.6 mg/kg (ip)	6,12,24,48 h		6,12,24,48 h
Disopyramide	Methylcellulose	Safe	20 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Endotoxin	Saline	High	8 mg/kg (ip)	6,12,24,48 h		
Endotoxin	Saline	Low	0.8 mg/kg (ip)	6,12,24,48 h		
Erythromycin Estolate (EE)	Methylcellulose	High	800 mg/kg/d	1,3,7,14 d		
Erythromycin Estolate (EE)	Methylcellulose	Low	80 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Ethanol	Corn Oil	High	15 g/kg/d	1,3,7,14 d		
Ethanol	Corn Oil	Low	1.5 g/kg/d	1,3,7,14 d		1,3,7,14 d
Ethinylestradiol	Corn Oil	High	15 mg/kg/d	1,3,7,14 d		
Ethinylestradiol	Corn Oil	Low	1.5 mg/kg/d	1,3,7,14 d		
Ethionine	Methylcellulose	High	200 mg/kg/d	1,3,7,14 d	1,3 d	
Ethionine	Methylcellulose	Low	20 mg/kg/d	1,3,7,14 d		
Ethosuximide	Methylcellulose	Safe	100 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Ethylenethiourea	Saline	High	920 mg/kg (po)	6,12,24,48 h		
Ethylenethiourea	Saline	Low	92 mg/kg (po)	6,12,24,48 h		
Fenarimol	Corn Oil	High	62.5 mg/kg/d	1,3,7,14 d		
Fenarimol	Corn Oil	Low	6.25 mg/kg/d	1,3,7,14 d		
Fenbendazole	Methylcellulose	High	3000 mg/kg/d	1,3,7,14 d		
Fenbendazole	Methylcellulose	Low	300 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Furan	Corn Oil	High	15 mg/kg/d	1,3,7,14 d		
Furan	Corn Oil	Low	1.5 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Hydrazine	Methylcellulose	High	100 mg/kg/d	1,3,7,14 d		
Hydrazine	Methylcellulose	Low	10 mg/kg/d	1,3,7,14 d		
Ibuprofen	Methylcellulose	Safe	94 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Ketoconazole	Methylcellulose	High	100 mg/kg/d	1,3,7,14 d		
Ketoconazole	Methylcellulose	Low	10 mg/kg/d	1,3,7,14 d		1,3 d
Lead nitrate	Saline	High	33 g/kg (iv)	1,3,7,14 d		
Lead nitrate	Saline	Low	3.3 g/kg (iv)	1,3,7,14 d		
Methapyrilene	Methylcellulose	High	60 mg/kg/d	1,3,7,14 d		
Methapyrilene	Methylcellulose	Low	6 mg/kg/d	1,3,7,14 d		
Methionine-choline deficient diet	In feed	High	60 g/kg/d	1,3,7,14 d		
Methyl Carbamate	Methylcellulose	High	400 mg/kg/d	1,3,7,14 d		

Methyl Carbamate	Methylcellulose	Low	40 mg/kg/d	1,3,7,14 d		
Methylcellulose		Control		1,3,7,14 d		1,3,7,14 d
Methylcellulose (ip)		Control		6,12,24,48 h		
Methylenedianiline	Corn Oil	High	50 mg/kg/d	1,3,7,14 d		
Methylenedianiline	Corn Oil	Low	5 mg/kg/d	1,3,7,14 d		
Methyleugenol	Methylcellulose	High	1000 mg/kg/d	1,3,7,14 d		
Methyleugenol	Methylcellulose	Low	100 mg/kg/d	1,3,7,14 d		
Methyl-tert-butyl ether	Corn Oil	High	1500 mg/kg/d	1,3,7,14 d		
Methyl-tert-butyl ether	Corn Oil	Low	150 mg/kg/d	1,3,7,14 d		
Microcystin-LR	Saline	High	20 µg/kg (iv)	6,12,24,48 h		
Microcystin-LR	Saline	Low	2 µg/kg (iv)	6,12,24,48 h		
Mirex	Corn Oil	High	10 mg/kg/d	1,3,7,14 d		
Mirex	Corn Oil	Low	1 mg/kg/d	1,3,7,14 d		
Molybdenum	Methylcellulose	High	500 mg/kg/d	1,3,7,14 d	3 d	
Molybdenum	Methylcellulose	Low	50 mg/kg/d	1,3,7,14 d		
Monocrotaline	H <sub>2</sub> O	High	160 mg/kg/d	1,3,7,14 d	1,3 d	
Monocrotaline	H <sub>2</sub> O	Low	16 mg/kg/d	1,3,7,14 d		
N-diethylnitrosamine	Methylcellulose	High	12 mg/kg/d	1,3,7,14 d		
N-diethylnitrosamine	Methylcellulose	Low	1.2 mg/kg/d	1,3,7,14 d		
Nifedipine	Methylcellulose	Safe	3 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Nitrofurantoin	Saline	High	150 mg/kg (ip)	1,3,7,14 d		
Nitrofurantoin	Saline	Low	15 mg/kg (ip)	1,3,7,14 d		1,3,7,14 d
Nitrosodiethanolamine	Methylcellulose	High	200 mg/kg/d	1,3,7,14 d		
Nitrosodiethanolamine	Methylcellulose	Low	20 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Nitrosomethylethylamine	Methylcellulose	High	75 mg/kg/d	1,3,7,14 d	1,3 d	
Nitrosomethylethylamine	Methylcellulose	Low	7.5 mg/kg/d	1,3,7,14 d	3,7,14 d	
N-nitrosodibutylamine	Methylcellulose	High	25 mg/kg/d	1,3,7,14 d		
N-nitrosodibutylamine	Methylcellulose	Low	2.5 mg/kg/d	1,3,7,14 d		1,3,7,14 d
N-nitrosomorpholine	Methylcellulose	High	35 mg/kg/d	1,3,7 d	1,3,7,14 d	
N-nitrosomorpholine	Methylcellulose	Low	3.5 mg/kg/d	1,3,7,14 d		
N-Nitrosopiperidine	Methylcellulose	High	200 mg/kg/d	1,3,7,14 d	1,3 d	
N-Nitrosopiperidine	Methylcellulose	Low	20 mg/kg/d	1,3,7,14 d		
NNK	Methylcellulose	High	20 mg/kg/d	1,3,7,14 d		
NNK	Methylcellulose	Low	2 mg/kg/d	1,3,7,14 d		
Pentachlorophenol	Methylcellulose	High	50 mg/kg/d	1,3,7,14 d		



Pentachlorophenol	Methylcellulose	Low	5 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Phenobarbital	Methylcellulose	High	80 mg/kg/d	1,3,7,14 d		
Phenobarbital	Methylcellulose	Low	8 mg/kg/d	1,3,7,14 d		
Piperonyl Butoxide	Methylcellulose	High	1200 mg/kg/d	1,3,7,14 d		
Piperonyl Butoxide	Methylcellulose	Low	120 mg/kg/d	1,3,7,14 d		
Potassium bichromate	Methylcellulose	High	10 mg/kg (ip)	6,12,24,48 h		
Potassium bichromate	Methylcellulose	Low	1 mg/kg (ip)	6,12,24,48 h		
Prazosin	Methylcellulose	Safe	1 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Pregnenolone-16 $\alpha$ -Carbonitrile	Methylcellulose	High	100 mg/kg/d	1,3,7,14 d		
Pregnenolone-16 $\alpha$ -Carbonitrile	Methylcellulose	Low	10 mg/kg/d	1,3,7,14 d		
Propranolol	Methylcellulose	Safe	40 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Pyridine	Methylcellulose	High	300 mg/kg/d	1,3,7,14 d	14 d	
Pyridine	Methylcellulose	Low	30 mg/kg/d	1,3,7,14 d		
Ranitidine	Methylcellulose	Safe	5 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Rifampicin	Methylcellulose	High	250 mg/kg/d	1,3,7,14 d		
Rifampicin	Methylcellulose	Low	25 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Saline (ip)		Control		6,12,24,48 h		6,12,24,48 h
Saline (ip)		Control		1,3,7,14 d		1,3,7,14 d
Saline (iv)		Control		6,12,24,48 h		6,12,24,48 h
Saline (iv)		Control		1,2,3,4 d		1,2,3,4 d
Saline (iv)		Control		1,3,7,14 d		1,3,7,14 d
Saline (po)		Control		6,12,24,48 h		6,12,24,48 h
Terfenadine	Methylcellulose	Safe	10 mg/kg/d	1,3,7,14 d		1,3,7,14 d
Thioacetamide	Saline	High	200 mg/kg (ip)	6,12,24,48 h	12,24,48 h	
Thioacetamide	Saline	Low	20 mg/kg (ip)	6,12,24,48 h	6,12,24,48 h	
Thioacetamide	Methylcellulose	High	19.2 mg/kg/d	1,3,7,14 d		
Thioacetamide	Methylcellulose	Low	1.92 mg/kg/d	1,3,7,14 d		
Thiobenzamide	In Feed	High	<i>ad libitum</i> 1 g/kg	1,3,7,14 d	1,3,7,14 d	
Thiobenzamide	In Feed	Low	<i>ad libitum</i> 0.1 g/kg	1,3,7,14 d		
Untreated		Control		1,3,7,14 d		1,3,7,14 d
Untreated		Control		6,12,24,48 h		6,12,24,48 h
Untreated		Control		1,2,3,4 d		1,2,3,4 d
WY-14643	Methylcellulose	High	60 mg/kg/d	1,3,7,14 d		

WY-14643	Methylcellulose	Low	6 mg/kg/d	1,3,7,14 d		
Xylidine	Corn Oil	High	600 mg/kg/d	1,3,7,14 d		
Xylidine	Corn Oil	Low	60 mg/kg/d	1,3,7,14 d		

### Example 2: Induction of Zone 3 Necrosis *in vitro*

Over 100 compounds were chosen based on evidence that they elicit one of eleven selected histopathology subtypes. The criteria for inclusion of a compound into the zone 3 necrosis mode was necrosis or individual cell necrosis in centrilobular zone with an increase in some or all serum transaminases. The compounds included in the zone 3 necrosis mode can be seen in Table 2. Rat primary hepatocytes were treated with the same compounds used in the *in vivo* experiments. In all cases, compounds were run in any given two-day period with an untreated and DMSO vehicle as negative controls. Each compound was delivered daily at a reference dose derived from the literature (when available), a 10x high dose, and 1/10 and 1/100 low doses for up to 72 hours. Three hepatocyte cultures for each dose were randomly assigned for harvest at 6, 24, 48 and 72 hours. In order to best identify genes characteristic of zone 3 necrosis, total RNA for all hepatocyte cultures from a given dose time point were pooled and converted to mRNA and cDNA for GeneCalling®. In GeneCalling, the cDNA is cut with a battery of restriction enzyme pairs in different combinations followed by amplification by PCR using specific primers linked to specific adaptors. After gel electrophoresis, the resulting fragments are identified based on the inherent information in the cDNA fragment: The flanking restriction site sequences on the ends, the size of the fragment and the species (and sometimes the tissue) origin of the DNA. This information is used to query public and proprietary databases. The fragments that do not match any sequences in the database are isolated, sequenced and identified as novel.

**Table 2: Zone 3 Necrosis *in vitro***

Compound	Annotation	Concentration (microM)	Zone 3 Necrosis Pathology
1,3-Dibromobenzene		63	
1,3-Dibromobenzene		250	Present
2AAF		10	
2AAF		50	
2-nitrofluorene		30	

2-nitrofluorene		150	
Acetamide		500	
Acetamide		2000	
Acetaminophen		100	
Acetaminophen		1000	Present
Aflatoxin		0.008	
Aflatoxin		0.04	
Allyl alcohol		16	
Allyl alcohol		80	
Allylformate		0.06	
Allylformate		0.3	
Amiodarone		5	
Amiodarone		20	
ANIT		3.1	
ANIT		13	
Aspirin	non-toxic	55.5	Absent
Aspirin	non-toxic	555	
Atorvastatin	non-toxic	0.0095	Absent
Atorvastatin	non-toxic	0.095	
Azaserine		100	
Azaserine		500	Present
Azobenzene		1.6	Absent
Azobenzene		8	
BCNU		50	
BCNU		250	
Butylhydroxytoluene		75	
Butylhydroxytoluene		150	
Bretylium	non-toxic	1.2	Absent
Bretylium	non-toxic	12	
Bromobenzene		600	
Bromobenzene		3000	Present
Carbamate		300	
Carbamate		1500	
Carbamazepine		200	
Carbamazepine		1000	
CCNU		8	
CCNU		40	
CdCl		0.1	
CdCl		0.5	
CeCl3		4	
CeCl3		20	Present
Cefuroxime	non-toxic	224	Absent
Cefuroxime	non-toxic	2240	
Chlordane		8	Absent
Chlordane		40	

CIDirect		1	
CIDirect		5	
Ciprofibrate		100	
Ciprofibrate		500	
Clofibrate		100	Absent
Clofibrate		400	
Clonidine	non-toxic	0.0165	Absent
Clonidine	non-toxic	0.165	
Colchicine		500	
Colchicine		2000	
Concanavalin A		5000	
Concanavalin A		20000	
Coumarin		63	
Coumarin		250	Present
4-chlorobenzotrifluoride		250	
4-chlorobenzotrifluoride		1000	
Cyclosporine		3.1	Absent
Cyclosporine		13	
Cyproterone		10	
Cyproterone		50	
2,4-diaminotoluene		0.8	
2,4-diaminotoluene		4	
Di(2-Ethylhexyl)phthalate		500	
Di(2-Ethylhexyl)phthalate		2000	
Dehydroepiandrosterone		1.2	
Dehydroepiandrosterone		6	
Deoxycholate		6.3	
Deoxycholate		25	
Dexamethasone		100	Absent
Dexamethasone		500	
Dibutyltin		0.2	
Dibutyltin		1	
Dichlorobenzene		100	Absent
Dichlorobenzene		500	
Diethylnitrosamine		60	
Diethylnitrosamine		300	Present
Diethylstilbestrol		5	
Diethylstilbestrol		50	
Dimethylnitrosamine		200	
Dimethylnitrosamine		1000	Present
Disopyramide	non-toxic	3	Absent
Disopyramide	non-toxic	30	
Dimethylformamide		1000	
Dimethylformamide		5000	Present
DMSO	control	0	Absent

Doxorubicin	non-toxic	0.5	Absent
Doxorubicin	non-toxic	5	
Endotoxin		30	
Endotoxin		100	
Erythromycin		20	Absent
Erythromycin		100	
Ethanol		200	Absent
Ethanol		1000	
Ethinylestradiol		25	
Ethinylestradiol		100	
Ethionine		200	
Ethionine		1000	Present
Ethosuximide	non-toxic	1000	Absent
Ethosuximide	non-toxic	10000	
Ethylenethiourea		200	
Ethylenethiourea		1000	
Fenarimol		20	Absent
Fenarimol		100	
Fenbendazole		16	
Fenbendazole		63	
Fluconazole	non-toxic	0.816	Absent
Fluconazole	non-toxic	8.16	
Gabapentin	non-toxic	2	Absent
Gabapentin	non-toxic	20	
Galactosamine		12	
Galactosamine		60	
Hydrazine		20	
Hydrazine		100	
Ibuprofen	non-toxic	50	Absent
Ketoconazole		2	Absent
Ketoconazole		10	
Mephénytoin	non-toxic	14.2	Absent
Mephénytoin	non-toxic	142	
Methapyriline		30	
Methapyriline		100	
Methylcholanthrene		40	Absent
Methylcholanthrene		200	
Methylenedianiline		1.4	
Methylenedianiline		7.8	
Methyleugenol		100	
Methyleugenol		500	
Microcystin		0.005	
Microcystin		0.025	
Minoxidil	non-toxic	0.166	Absent
Minoxidil	non-toxic	1.66	

Mirex		50	Absent
Mirex		100	
Molybdenum		20	
Molybdenum		50	Present
Monocrotaline		30	
Monocrotaline		100	Present
Methyl-tert-butyl ether		1000	
Methyl-tert-butyl ether		4000	
Nifedipine	non-toxic	0.335	Absent
Nifedipine	non-toxic	3.35	
Nitrofurantoin		4	Absent
Nitrofurantoin		20	
Nitrosodibutylamine		200	Absent
Nitrosodibutylamine		1000	
Nitrosodiethanolamine		1000	Absent
Nitrosodiethanolamine		5000	
Nitrosomethylethylamine		200	
Nitrosomethylethylamine		1000	Present
Nitrosomorpholine		750	
Nitrosomorpholine		3750	Present
Nitrosopiperidine		640	
Nitrosopiperidine		3200	Present
NNK		200	
NNK		1000	
Norethindrone		40	Absent
Norethindrone		200	
Pentachlorophenol		19	Absent
Pentachlorophenol		38	
Piperonyl		20	
Piperonyl		100	
Prazosin	non-toxic	0.0148	Absent
Prazosin	non-toxic	0.148	
Pregnenolone		38	Absent
Pregnenolone		150	
Propranolol	non-toxic	0.125	Absent
Propranolol	non-toxic	1.25	
Pyridine		800	
Pyridine		4000	Present
Ranitidine	non-toxic	0.128	Absent
Ranitidine	non-toxic	1.28	
Rifampicin		20	Absent
Rifampicin		100	
Terfenadine	non-toxic	0.15	Absent
Terfenadine	non-toxic	1.5	
Thioacetamide		500	

Thioacetamide		2000	Present
Thiobenzamide		8	
Thiobenzamide		40	Present
Untreated	control	0	Absent
Verapamil	non-toxic	0.1	Absent
Verapamil	non-toxic	1	
WY14643		20	
WY14643		100	
Xylidine		13	
Xylidine		50	

### EXAMPLE 3: METHODS OF ANALYSIS

#### Data Preparation:

We used GeneCalling® to estimate the activity of several thousand transcripts simultaneously. These data generally have ~5% missing data and are log normally distributed.

5 The data are log transformed and missing values are filled using k-nearest neighbor (knn) replacement [14]. The knn algorithm was initially validated using a complete data set and randomly eliminating constant percentages of the data. It was determined that using correlation as a similarity index and imputing missing values with 6 nearest neighbors resulted in the smallest error of prediction.

10 Initially our data sets contained between 6000 and 8000 genes, which poses two problems. These large numbers of genes make most marker selection procedures computationally intractable with most computer algorithms. Second, the inclusion of markers with low variation, or low association with pathology results in a significant risk of choosing markers that over fit the models. To eliminate these problems we imposed an initial filter on the

15 data, requiring that there be a significant difference between negative control samples and positive control samples for each pathology mode (Kruskal-Wallis test,  $p < 0.001$ ). Depending on mode, this process reduced our gene set to a more tractable number of genes (approximately 200-800 genes depending on mode).

#### Initial Marker Selection:

20 We define a marker as a gene that helps to explain some variation in pathology. In order to avoid selecting markers that particularly fit our current data set well at the expense of predictability outside our training set, we employed a leave one out cross-validation method to

identify markers that contribute some explanatory power to the data set. Specifically, after the Kruskal-Wallis filter, a series of leave one out models are created leaving out all of the samples for each compound until all compounds have been left out once. This process results in a marker list and a count of the number of leave-one out models the marker was used in. This marker list contains all of the genes that explain some portion of the variation in pathology but is almost certain to over fit the data because of its size. In order to refine this marker list a second series of leave one compound out models is created for each different count of markers within the marker list. For example, the initial leave one out model may produce a gene list consisting of 5 genes that occur at frequencies of 20, 19, 19, 2, and 1 leave one out models respectively. The first step will use all genes that occur in 1 or more leave one out models, the second step 2 or more, the third 19 or more and finally 20 or more. The genes that are considered to be markers will have been used in a majority of models and result in a highly sensitive model. In most cases the first modeling step resulted in a sharp cutoff (e.g. 19 or more in the above example), which guides the marker selection process. In a few cases, the change in frequency was so gradual that no clear cutoff was available. Marker selection then proceeded with the most sensitive model first, and then the most specific model and in the case of ties the least number of markers.

### **Models:**

The above process is a general strategy that is applied to all of our marker selection models. The models we used covered a range of statistical power and assumption stringency. The most powerful model with the strongest assumptions is a linear discriminant analysis, followed by logistic regression and finally by classification trees, which is virtually devoid of assumptions but does have a cost in terms of predictivity. These modeling methods are common statistical procedures that need not be developed here [15] [16] for a more detailed discussion). All three methods went through the algorithm outlined above with the exception of discriminant analysis, which did not utilize a Kruskal-Wallis filter. Both discriminant analysis and logistic regression create poor models when too many intercorrelated variables are used. To minimize this problem, these methods utilized a stepwise selection procedure (incorporating both forward and reverse selection) to select the best discriminating set of markers.



### ***In vivo* Pathology Annotation:**

In order to construct *in vivo* models, the pathology of each sample had to be determined. Pathology was assigned to each liver sample by the pathologists at Bayer's Stillwell, Kansas facility. Criteria for inclusion of a compound into a particular mode included the following:

- 5           *i. Zone-3 Necrosis:* Necrosis or individual cell necrosis in centrilobular zone with an increase in some or all serum transaminases.
- ii. Cholestasis:* Increased plasma bilirubin with bile duct necrosis or hyperplasia.
- iii. Hypertrophy:* Increase in cell size and liver weight.
- iv. Genotoxic Carcinogens:* Some evidence of mild parenchymal damage *in vivo*  
10           which may be associated with an increase in mitosis. Compounds assigned to this group must have strong historical documentation. Compounds from other pathology modes cannot be added to this list.
- v. Non-genotoxic Carcinogens:* Evidence of parenchymal changes and an increase in mitosis. Compounds assigned to this group must have strong historical  
15           documentation. Compounds from other pathology modes cannot be added to this list.
- vi. Steatosis:* Increase in lipid accumulation or "vacuolar degeneration."
- vii. Zone 1 Necrosis:* Necrosis or individual cell necrosis in the periportal zone, with an increase in some or all serum transaminases.
- 20           *vii. Inflammation:* Increase in inflammatory cells (e.g. Kupffer cells, neutrophils, macrophages, lymphocytes)
- viii. Apoptosis:* Shrinking or fragmentation of the nucleus and increased "blebbing."

### ***In vitro* Pathology Annotation:**

25           In order to construct *in vitro* models we need to make a decision about what pathology each sample represents. Two approaches were used. The first is to ascribe the *in vivo* compound annotation to the two highest concentrations *in vitro* (these are traditionally within five fold of

each other). For example, clofibrate produced hypertrophy *in vivo*, so the two highest doses *in vitro* are used as a positive control for the model construction process. The second annotation strategy uses a nearest neighbor algorithm to assign annotation from *in vivo* samples to *in vitro* samples. Briefly, each *in vitro* sample was correlated, across *in vivo* markers, to all *in vivo* samples of the same compound. The annotation of the most correlated *in vivo* sample was used as the *in vitro* annotation.

### Final Marker Selection:

This process resulted in 6 sets of models being generated for each mode of pathology (LDA, logistic, and classification trees for each annotation strategy, nearest neighbor and high dose). These markers are then correlated with the original data set (between 6000 and 8000 genes) and additional correlated markers ( $r \geq 0.60$  across 329 samples, up to 3 per marker) were added back in to the final gene set for representation on a microarray.

### Example 4: Identification of Zone 3 Necrosis Related Genes *in vivo*

SP = Secreted Protein  
NC = Novel Rat Composition  
NU = Novel Rat Utility

**Table 3**

	ACCN O	TOX MARKER ASSIGNMEN T	SEQ ID NO:	Definition	Description	Bin
NU	scr_gb- af03887 0_4	1	1	Rattus norvegicus betaine homocysteine methyltransferase (BHMT) [AF038870].	Betaine-homocysteine methyltransferase (BHMT) catalyzes the transfer of an N-methyl group from betaine to homocysteine to produce dimethylglycine and methionine, respectively. The enzyme is found in the pathway of choline oxidation and is abundantly expressed in liver and kidney. It has been known for at least 50 years that alterations in methionine metabolism occur in human liver cirrhosis. Recently human BHMT had been shown to be a zinc metalloenzyme [14] [15].	Amino Acid Metabolism

NC NU	scr_gb- z83053_ 3	2	2	Rat gene fragment - 1984 bp. 88% SI (1241/1396) to Mus musculus betaine- homocysteine methyltransferase 2 (Bhmt2) [AF257474].	Betaine-homocysteine methyltransferase (BHMT) catalyzes the transfer of an N-methyl group from betaine to homocysteine to produce dimethylglycine and methionine, respectively. The enzyme is found in the pathway of choline oxidation and is abundantly expressed in liver and kidney. It has been known for at least 50 years that alterations in methionine metabolism occur in human liver cirrhosis. Recently human BHMT had been shown to be a zinc metalloenzyme [14] [15].	Amino Acid Metabolism
NU	scr_gb- x95189_ 4	3	3	Rattus norvegicus Trihydroxycoprostanoyl-CoA Oxidase [X95189].	Rat liver peroxisomes contain three acyl-CoA oxidases: palmitoyl-CoA oxidase, pristanoyl-CoA oxidase, and trihydroxycoprostanoyl-CoA oxidase. Mammalian liver peroxisomes are capable of beta-oxidizing a variety of substrates including very long chain fatty acids and the side chains of the bile acid intermediates di- and trihydroxycoprostanic acid. The first enzyme of peroxisomal beta-oxidation is acyl-CoA oxidase [16].	Lipid Metabolism
NU	scr_gb- m59814_ 4	4	4	Rattus norvegicus Ephrin type-B receptor 1 precursor (EphB1) [P09759]	Eph receptor tyrosine kinases and their membrane-bound ligands, ephrins, have thus emerged as mediators of cell-contact- dependent repulsion. The actin cytoskeleton is also a major target of the intracellular pathways activated by Eph receptors [17]. More specifically, activation of EphB1 by its ligand, ephrin-B1/Fc has been shown to recruit Nck to native receptor complexes and activate c-Jun kinase (JNK/SAPK) [18]	Cell Cycle Regulation (Regulation Of Proliferation)
NU	scr_gb- m29358 _5	5	5	Rattus norvegicus ribosomal protein S6 [M29358].	It has been known for 20 years that the ribosomal protein S6 is rapidly phosphorylated when cells are stimulated to grow or divide [19]. S6 is phosphorylated in response to mitogens by activation of one or more protein kinase cascades. Members of the 90 kDa S6 kinases are activated in vitro by 42 kDa and 44 kDa MAP kinases, which are in turn activated by mitogen-dependent activators [20].	Protein Metabolism
NU	aj29773 6	6	6	Rattus norvegicus heat shock protein 86 (hsp86) [AJ428213].	In addition to appearing in response to biological stresses, heat shock proteins are expressed as 'chaperones' by some cells living in physiological conditions. Among these proteins, the Hsp90 family, consisting of isoforms Hsp84 and Hsp86, seems to function under normal growth conditions in the pathways of numerous signal transducers, cell cycle and developmental regulators. [21, 22]	Protein Metabolism
NU	j00719	7	7	Rattus norvegicus cytochrome p-450 isoform, (phenobarbital- inducible or 2B1) [J00719] [P04167].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [23].	Detoxification Response / Biotransformation -TOX

NU	j00720	8	8	Rattus norvegicus cytochrome p-450 isoform, (phenobarbital-inducible or 2B2) [P04167] [J00719].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [23].	Detoxification Response / Biotransformation -TOX
NU	j00728	9	9	Rattus norvegicus cytochrome p-450 isoform, (phenobarbital-inducible, 2B1, or 2B2) [P00176] [P04167] [Q64584].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [23].	Detoxification Response / Biotransformation -TOX
NU	i00320	10	10	Rattus norvegicus cytochrome p-450 isoform (phenobarbital-inducible, 2B1, or 2B2) [J00719] [P00176] [P04167].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [23].	Detoxification Response / Biotransformation -TOX
NU	m11251	11	11	Rattus norvegicus cytochrome p-450 isoform (phenobarbital-inducible, 2B1, or 2B2) [Q64584] [P00176] [P04167].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [23].	Detoxification Response / Biotransformation -TOX
NU	m26125	12	12	Rattus norvegicus epoxide hydrolase [M26125].	Epoxide formation from drugs, chemicals, food additives and environmental pollutants is catalyzed by cytochrome P-450 dependent monooxygenase(s). Epoxides are converted to glycols or dihydrodiols by epoxide hydrolase. These enzymes are known to be present in the microsomes of different mammalian tissues and in the hepatic nuclei from rats and humans. The balance between the epoxide forming (AHH) and metabolizing (EH) enzyme activities may provide information about the "epoxide exposure" of a tissue [Kuklin, 1976 #2].	Detoxification Response / Biotransformation -TOX

NU	m34452	13	13	Rattus norvegicus cytochrome P450e-L (P450IIB2) [M34452].	Epoxide formation from drugs, chemicals, food additives and environmental pollutants is catalyzed by cytochrome P-450 dependent monooxygenase(s). Epoxides are converted to glycols or dihydrodiols by epoxide hydrolase. These enzymes are known to be present in the microsomes of different mammalian tissues and in the hepatic nuclei from rats and humans. The balance between the epoxide forming (AHH) and metabolizing (EH) enzyme activities may provide information about the "epoxide exposure" of a tissue [Kuklin, 1976 #2].	Detoxification Response / Biotransformation -TOX
NU	u33546	14	14	Rattus norvegicus CYP2B16P [CAB35441].	CYP2B16P is an apparent pseudogene in the rat cytochrome P450 2B (CYP2B) subfamily [24].	Detoxification Response / Biotransformation -TOX
NU	x74673	15	15	Rattus norvegicus aflatoxin B1 aldehyde reductase (AFAR) [X74673].	Aflatoxin B1 aldehyde reductase/Succinic semialdehyde reductase is believed to be involved in the detoxification of xenobiotic carbonyl compounds [25].	Detoxification Response / Biotransformation -TOX
SP	scr_gb-x13044_4	16	16	Rattus norvegicus MHC-associated invariant chain gamma (Ia antigen-associated invariant chain) (Ii) [X13044].	MHC class II molecules at the surface of antigen presenting cells present antigenic peptides to CD4+ T helper cells. Ii plays a critical role in MHC class II antigen processing by stabilizing peptide-free class II alpha/beta heterodimers [26].	Immunity And Defense
SP	scr_gb-x14254_5	17	17	Rattus norvegicus MHC-associated invariant chain gamma (Ia antigen-associated invariant chain) (Ii) [X14254].	MHC class II molecules at the surface of antigen presenting cells present antigenic peptides to CD4+ T helper cells. Ii plays a critical role in MHC class II antigen processing by stabilizing peptide-free class II alpha/beta heterodimers [26].	Immunity And Defense
NU	scr_gb-bi27563_8_1	18	18	Rattus norvegicus Ras-related protein Rab-2 [P05712].	Rab proteins form the largest branch of the Ras superfamily of GTPases. They are localized to the cytoplasmic face of organelles and vesicles involved in the biosynthetic/secretory and endocytic pathways in eukaryotic cells [27].	Intracellular Transport
	scr_gb-x66871_3	19	19	Rattus norvegicus calpactin I heavy chain (annexin II) [X66871].	The annexins are a family of proteins that bind acidic phospholipids in the presence of Ca <sup>2+</sup> . Because annexin II bridge secretory granules to plasma membrane it has suggested that this protein may play a role in Ca(2+)-dependent exocytosis. Annexin II tetramer has also been demonstrated on the extracellular face of some metastatic cells where it mediates the binding of certain metastatic cells to normal cells. Annexin II tetramer is a major cellular substrate of protein kinase C and pp60src [28].	Intracellular Transport
NU	scr_gb-l49379_3	20	20	Rattus norvegicus canalicular multispecific organic anion transporter (cMOAT) [L49379].	cMOAT mediates the hepatobiliary excretion of numerous organic anions. It has been shown that both multidrug resistance-associated protein (MRP1) and canalicular multispecific organic anion transporter (cMOAT/MRP2) have the ability to extrude glutathione conjugates (GS-X pump activity) from cells [29] [30].	Oxidative Stress-TOX

SP	scr_sc-132690501_1	21	21	Rat gene fragment - 775 bp. 85% SI (618/722) to Homo sapiens inter-alpha-trypsin inhibitor heavy chain ITIH1 [X63652].	Inter-alpha-trypsin inhibitor (ITI) is a complex protein containing two heavy polypeptide chains (H1 and H2) and a light chain, which in the free state is known as bikunin [31]. ITI is a 220 kDa serine proteinase inhibitor found in human serum [32].	Other
NC	scr_gb-aw141735_3	22	22	Rat gene fragment - 1561 bp. 98% SI (1002/1022) to Mus musculus serine proteinase inhibitor mBM2A [U96701].	Serine proteinase inhibitors (serpins) are classically regulators of extracellular proteolysis. Evidence suggests that some function intracellularly as well [33].	Other
SP	af184983	23	23	Rattus norvegicus osteoactivin [AF184983] .	Osteoactivin cDNA was recently isolated from long bone and calvaria . In primary rat osteoblast cultures it was expressed at the highest levels during the later stages of matrix maturation and mineralization and correlated with the expression of alkaline phosphatase and osteocalcin. [34].	Unknown
NU	scr_cg-22510674_1	24	24	Unknown, 241 bp.		Novel
NU	scr_cg-57215224_1	25	25	Unknown, 283 bp.		Novel
NU	scr_gb-aa850767_2	26	26	Unknown, 642 bp.		Novel
NU	scr_gb-ai011994_2	27	27	Unknown, 866 bp.		Novel
NU	scr_gb-aw142293_1	28	28	Unknown, 629 bp.		Novel
NU	scr_gb-bm383327_1	29	29	Unknown, 1145 bp.		Novel
NU	scr_gb-bm386625_1	30	30	Unknown, 3087 bp.		Novel
NU	scr_sc-133556969_1	31	31	Unknown, 434 bp.		Novel
NU	scr_sc-170142736_1	32	32	Unknown, 221 bp.		Novel
NU	scr_sc-2563586_2	33	33	Unknown, 581 bp.		Novel
NU	scr_sc-87618257_1	34	34	Unknown, 221 bp.		Novel

**Example 5: Identification of Zone 3 Necrosis Related Genes *in vitro***

SP = Secreted Protein  
 NC = Novel Rat Composition  
 NU = Novel Rat Utility

**Table 4**

5

	ACCNO	TOX MARKER ASSIGN MENT	SEQ ID NO	Definition	Description	Bin
NU	cszr_96561 134_837604 93	35	35	Rattus norvegicus Carbamoyl-phosphate synthase [ammonia] (CPSASE I), mitochondrial precursor [P07756].	Mitochondrial protein involved in the urea acid cycle of ureotelic animals where the enzyme plays an important role in removing excess ammonia from the cell. Catalytic Activity: 2 ATP + NH(3) + CO(2) + H(2)O = 2 ADP + ORTHOPHOSPHATE + CARBAMOYL PHOSPHATE [17].	Amino Acid Metabolism
NU	scr_gb- x83855_1	36	36	Rattus norvegicus hepatocyte EP3alpha receptor [X83855].	EP3 receptors for Prostaglandin (PG) E(2) are primarily involved in inhibition of adenylyl cyclase via G(i) activation, and in Ca(2+)-mobilization through Gbetagamma from G(i). Along with G(i) activation, the EP3 receptor can stimulate cAMP production via G(s) activation [18].	Carbohydrate Metabolism
NU	cszr_229800 465_190907 286	37	37	Rattus norvegicus non-neuronal enolase (NNE) (alpha-alpha enolase, 2-phospho-D-glycerate hydrolase [X02610].	Enolase is a vital enzyme of the glycolytic pathway. It exists mainly in two forms, non-neuronal enolase (NNE) and neuron specific enolase (NSE). Catalytic Activity: 2-phospho-D-glycerate = phosphoenolpyruvate + H(2)O [19].	Carbohydrate Metabolism
SP	scr_gb- bi277612_1	38	38	Rat gene fragment - 1381 bp. 89% SI (816/910) to Mus musculus for beta-hexosaminidase [Y00964].	Two genes, HEXA and HEXB, encode the alpha- and beta-subunits, respectively, of human beta-hexosaminidase. In the mouse, the corresponding genes are termed Hexa and Hexb. The subunits have the capacity to degrade a variety of substrates including oligosaccharides, glycosaminoglycans, and glycolipids containing beta-linked N-acetylglucosaminyl or N-galactosaminyl residues [20].	Carbohydrate Metabolism
NU	scr_gb- j05266_3	39	39	Rattus norvegicus mitochondrial H+-ATP synthase alpha subunit [J05266].	H+-ATP synthase catalyzes the synthesis and/or hydrolysis of ATP [21].	Energy Metabolism
NU	scr_gb- m37394_5	40	40	Rattus norvegicus epidermal growth factor receptor (Egfr) [M37394].	Egfr is involved in the initiation of oncogenic effect such as DNA synthesis, enhanced cell growth, invasion, and metastasis. Specific abrogation of EGFR results in cell cycle arrest, apoptosis, or dedifferentiation of cancer cells [22].	Cell Cycle Regulation (Regulation Of Proliferation)

NU	scr_gb-m64300_4	41	41	Rattus norvegicus extracellular signal-related kinase (ERK2) [M64300].	The Raf/MEK/ERK signaling was the first MAP kinase cascade to be characterized. It is probably one of the most well known signal transduction pathways among biologists because of its implication in a wide variety of cellular functions as diverse -and occasionally contradictory- as cell proliferation, cell-cycle arrest, terminal differentiation and apoptosis [23].	Cell Cycle Regulation (Regulation Of Proliferation)
SP	scr_gb-bi294409_1	42	42	Rat gene fragment - 526 bp. 88% SI (313/355) to Mus musculus type I interferon receptor soluble isoform precursor (IFNAR2) [AF013486].	IFNAR-2, is expressed ubiquitously, and exists as both transmembrane and soluble forms. Recent evidence suggests murine IFNAR-2 as an efficient regulator of IFN responses. Type I interferons are cytokines that are important in defense against viral infections well as in the control of cell proliferation [24] [25].	Cell Cycle Regulation (Regulation Of Proliferation)
NU	scr_gb-ab015747_3	43	43	Rattus norvegicus interleukin-4 receptor (membrane-bound form) [AB015747].	IL-4 is a pleiotropic cytokine which plays a pivotal role in shaping immune responses. The effects of IL-4 are mediated after binding to high affinity receptor complexes present on hematopoietic as well as non-hematopoietic cells. There is also evidence that IL-4 interaction with its receptor leads to signal transduction mechanisms that result in cellular proliferation and / or gene activation [26].	Cell Cycle Regulation (Regulation Of Proliferation)
NU	scr_sc-191879433_1	44	44	Rattus norvegicus Crk-associated substrate, p130 [D29766].	The Crk-associated substrate (Cas) is a unique docking protein with a Src homology 3 (SH3) domain. Aberrant CAS tyrosine phosphorylation may contribute to cell transformation by certain oncoproteins, including v-Crk and v-Src, and to tumor growth and metastasis [27] [28].	Cell Cycle Regulation (Regulation Of Proliferation)
NC	scr_sc-140438096_1	45	45	Rat gene fragment - 383 bp. 98% SI (125/127) to Homo sapiens Diacylglycerol kinase, delta [Q16760].	Diacylglycerol kinase (DGK) plays an important role in the signal transduction through modulating the balance between two signaling lipids, diacylglycerol and phosphatidic acid. Diacylglycerol is a protein kinase c activator. Thus, DGK is considered to regulate protein kinase C activity through the reduction of diacylglycerol [29] [30].	Cell Cycle / Proliferation (Basic Machinery)
NU	scr_gb-x87157_5	46	46	Rattus norvegicus neurotensin endopeptidase [X87157].	Neurotensin is a 13-amino acid hormonal peptide which was first isolated from bovine hypothalamus. It is present in the digestive tract as well as in the central nervous system. It has a variety of biological activities as a central neurotransmitter or neuromodulator, and a peripheral hormone [20].	Cellular Communication
NU	scr_gb-u66707_2	47	47	Rattus norvegicus densin-180 [U66707].	Densin-180 is a transmembrane protein that is strongly associated with the postsynaptic density in CNS neurons and is believed to function as a synaptic adhesion molecule [31].	Cellular Communication



NU	scr_gb-af017393_2	48	48	Rattus norvegicus cytochrome P4502F4 (CYP4502F4) [AF017393].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [32].	Detoxification Response / Biotransformation -TOX
NU	scr_sc-134241980_1	49	49	Rattus norvegicus cytochrome P450 2B3 (CYP2B3) [U16214].	The cytochromes P-450 are among the major constituent proteins of the liver mixed function monooxygenases. They play a central role in the metabolism of steroids, the detoxification of drugs and xenobiotics, and the activation of procarcinogens. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics [32].	Detoxification Response / Biotransformation -TOX
NC	scr_sc-191609675_1	50	50	Rat gene fragment - 217 bp. 94% SI (67/71) to Streptococcus pneumoniae DNA polymerase III, alpha subunit R6 [AAK99055].	DNA polymerase III is a replicative enzyme known to be essential in the DNA synthesis of Gram-positive bacteria [33].	DNA Metabolism
NU	scr_gb-x17037_2	51	51	Rat OX40 antigen [X17037].	OX40 is a member of the tumor necrosis factor family which is expressed by activated T lymphocytes [34].	Immunity And Defense
SP	scr_gb-bi291805_1	52	52	Rat gene fragment - 528 bp. 76% SI (290/380) to Homo sapiens IgG Fc binding protein [D84239].	Fc gamma BP is widely expressed on mucosal surfaces and in external secretions lending support to the concept that Fc gamma BP is an important component of mucosal immunological defenses [35].	Immunity And Defense
NU	scr_gb-aj000696_5	53	53	Rattus norvegicus novel kinesin-related protein, KIF1D [AJ000696].	The proteins of the kinesin superfamily (KIFs) are microtubule-based molecular motors whose functions include the transport of membrane-bound organelles. The KIF1 subfamily members are monomeric and contain a number of amino acid inserts in surface loops [36].	Intracellular Transport
NU	scr_gb-d79221_3	54	54	Rattus norvegicus r-sly1 [U35364].	r-sly1 is a mammalian homologue to yeast Sly1p which plays a critical role in endoplasmic reticulum to Golgi apparatus vesicle trafficking [37].	Intracellular Transport

NU	m61937	55	55	Rattus norvegicus dihydrodiol dehydrogenase [M61937].	Dihydrodiol dehydrogenase(s) (DD) have been implicated in the detoxication of proximate (trans-dihydrodiol) and ultimate carcinogenic (anti-diol-epoxide) metabolites of polycyclic aromatic hydrocarbons (PAHs). Although this pathway suppresses the formation of the PAH anti- and syn-diol epoxides (ultimate carcinogens), the process of autoxidation is anticipated to yield reactive oxygen species (ROS) [38].	Oxidative Stress-TOX
NU	cszr_229602935_183895355	56	56	Rattus norvegicus metallothionein-i (mt-1) [J00750].	Metallothionein (MT) is a small, cysteine-rich, metal-binding protein. MT synthesis is induced by various stimuli such as heavy metals, oxidative stress, anticancer drugs and fasting stress. MT is capable of not only reducing metal toxicity but also scavenging free radicals [39].	Oxidative Stress-TOX
NU	scr_gb-af106944_3	57	57	Rattus norvegicus Peroxiredoxin III [AF106944].	Peroxiredoxins are novel family of anti-oxidative proteins comprise six members in mammals. They share a common reactive Cys residue in the N-terminal region, and are capable of serving as a peroxidase and involve thioredoxin and/or glutathione as the electron donor [40].	Oxidative Stress-TOX
NU	scr_gb-m11794_3	58	58	Rattus norvegicus metallothionein-2 and metallothionein-1 genes [M11794].	Metallothionein (MT) is a small, cysteine-rich, metal-binding protein. MT synthesis is induced by various stimuli such as heavy metals, oxidative stress, anticancer drugs and fasting stress. MT is capable of not only reducing metal toxicity but also scavenging free radicals [39].	Oxidative Stress-TOX
NU	scr_gb-af069306_1	59	59	Rattus norvegicus transaldolase [AF069306].	Transaldolase is a key enzyme of the reversible nonoxidative branch of the pentose phosphate pathway (PPP) that is responsible for the generation of NADPH to maintain glutathione at a reduced state (GSH) and, thus, to protect cellular integrity from reactive oxygen intermediates (ROIs) [41].	Oxidative Stress-TOX
NU	scr_gb-d17310_4	60	60	Rattus norvegicus steroid 3-alpha-dehydrogenase [D17310].	Steroid 3-alpha-dehydrogenase is an important multifunctional oxidoreductase capable of metabolizing steroid hormones, polycyclic aromatic hydrocarbons, and prostaglandins. It is also required for bile acid synthesis and has been suggested to play an important role in net bile acid transport across the hepatocyte [42].	Oxidative Stress-TOX
NC	scr_gb-bf281368_2	61	61	Rat gene fragment - 1086 bp. 80% SI (754/938) to Human Prt1 homolog [U62583].	PRT1 is a component of translation initiation factor eIF-3 and originally discovered in Saccharomyces cerevisiae [43].	Protein Metabolism
NU	scr_gb-u56407_3	62	62	Rattus norvegicus ubiquitin conjugating enzyme [U56407].	Ubiquitin-conjugating enzymes (UBC) catalyze the covalent attachment of ubiquitin to target proteins and are distinguished by the presence of a UBC domain required for catalysis [44].	Protein Metabolism

NC	scr_gb-ai406674_1	63	63	Rat gene fragment - 796 bp. 96% SI (634/660) Mus musculus heterogeneous nuclear ribonucleoprotein C, clone MGC:5715 IMAGE:3499283 [BC004706].	Heterogeneous nuclear ribonucleoprotein (hnRNP) complexes, the structures that contain heterogeneous nuclear RNA and its associated proteins, constitute one of the most abundant components of the eukaryotic nucleus. hnRNPs appear to play important roles in the processing, and possibly also in the transport, of mRNA [45].	RNA metabolism
NC	scr_gb-bf290678_2	64	64	Rat gene fragment - 716 bp. 84% SI (542/643) to Mus musculus heterogeneous nuclear ribonucleoprotein G, splice variant 1 [AJ237847].	Heterogeneous nuclear ribonucleoprotein (hnRNP) complexes, the structures that contain heterogeneous nuclear RNA and its associated proteins, constitute one of the most abundant components of the eukaryotic nucleus. hnRNPs appear to play important roles in the processing, and possibly also in the transport, of mRNA [45].	RNA metabolism
NC	scr_gb-bi288503_1	65	65	Rat gene fragment - 456 bp. 94% SI (430/456) to Mus musculus cardiac lineage protein 1 (Clp1) [AY090614].	Mouse Clp-1 is a potential cardiac transcriptional regulatory factor [46].	RNA Metabolism
NU	scr_gb-d86383_2	66	66	Rattus norvegicus Hex [D86383].	Hex is a homeobox protein which is believed to function as a transcriptional repressor and may be involved in the differentiation and/or maintenance of the differentiated state in hepatocytes [47].	RNA metabolism
NU	scr_sc-133366194_1	67	67	Rattus norvegicus thymosin beta-10 [M58405].	Thymosin beta 10 is one of a small family of proteins closely related in sequence to thymosin beta 4, recently identified as an actin-sequestering protein [48].	Structural Repair-TOX
NC	cszr_230290139_182026368	68	68	Rat gene fragment - 249 bp. 91% SI (227/248) to Mus musculus sex-determination protein homolog Fem1a [AF064447].	FEM-1 is a signal-transducing regulator in the C. elegans sex-determination pathway. The existence of FEM-1 homologs in the mouse raises the possibility that evolutionary conservation of ancient FEM-1 signaling interactions may play a role in vertebrate cell-fate determination [49].	Other
NU	scr_gb-ai013477_2	69	69	Rattus norvegicus VL30 element [M91234].	VL30 elements constitute a family of retrotransposons that are associated with cancer by their overexpression in rodent malignancies, their induction in a fibroblast response to anoxia which shares features with the malignant phenotype, and their presence recombined into Harvey murine sarcoma virus (HaSV) and Kirsten murine sarcoma virus [50].	Other
NU	scr_gb-m91235_3	70	70	Rattus norvegicus VL30 element [M91235].	VL30 elements constitute a family of retrotransposons that are associated with cancer by their overexpression in rodent malignancies, their induction in a fibroblast response to anoxia which shares features with the malignant phenotype, and their presence recombined into Harvey murine sarcoma virus (HaSV) and Kirsten murine sarcoma virus [50].	Other

NU	cszr_204152 648_191521 095	71	71	Unknown, 63 bp.	Novel
NU	cszr_204152 792_191517 979	72	72	Unknown, 133 bp.	Novel
NU	cszr_204229 614_191891 958	73	73	Unknown, 124 bp.	Novel
NU	cszr_204229 615_191892 510	74	74	Unknown, 124 bp.	Novel
NU	scr_gb- aa801331_1	75	75	Unknown, 1252 bp.	Novel
NU	scr_gb- aa899865_3	76	76	Unknown, 1241 bp.	Novel
NU	scr_gb- aa997629_1	77	77	Unknown, 396 bp.	Novel
NU	scr_gb- aa997691_1	78	78	Unknown, 473 bp.	Novel
NU	scr_gb- ai411514_4	79	79	Unknown, 1221 bp.	Novel
NU	scr_gb- aw142560_3	80	80	Unknown, 695 bp.	Novel
NU	scr_gb- aw533305_2	81	81	Unknown, 771 bp.	Novel
NU	scr_gb- aw915573_2	82	82	Unknown, 2262 bp.	Novel
NU	scr_gb- be108509_1	83	83	Unknown, 422 bp.	Novel
NU	scr_gb- be111483_1	84	84	Unknown, 445 bp.	Novel
NU	scr_gb- be120910_1	85	85	Unknown, 482 bp.	Novel
NU	scr_gb- bf285287_1	86	86	Unknown, 784 bp.	Novel
NU	scr_gb- bf390383_1	87	87	Unknown, 486 bp.	Novel
NU	scr_gb- bf558463_2	88	88	Unknown, 921 bp.	Novel
NU	scr_gb- bf560709_1	89	89	Unknown, 525 bp.	Novel
NU	scr_gb- bg662990_2	90	90	Unknown, 930 bp.	Novel
NU	scr_gb- bi278552_1	91	91	Unknown, 1060 bp.	Novel
NU	scr_gb- bi278749_1	92	92	Unknown, 1158 bp.	Novel
NU	scr_gb- bi295938_1	93	93	Unknown, 1241 bp.	Novel

NU	scr_gb- bi296376_1	94	94	Unknown, 2695 bp.		Novel
NU	scr_gb- bm384392_1	95	95	Unknown, 2423 bp.		Novel
NU	scr_gb- bm387477_1	96	96	Unknown, 610 bp.		Novel
NU	scr_gb- bm986259_1	97	97	Unknown, 1047 bp.		Novel
NU	scr_gb- s69874_5	98	98	Unknown, 1191 bp.		Novel
NU	scr_sc- 119263563_1	99	99	Unknown, 384 bp.		Novel
NU	scr_sc- 132556005_1	100	100	Unknown, 181 bp.		Novel
NU	scr_sc- 132570828_1	101	101	Unknown, 130 bp.		Novel
NU	scr_sc- 132947646_1	102	102	Unknown, 50 bp.		Novel
NU	scr_sc- 133387221_1	103	103	Unknown, 296 bp.		Novel
NU	scr_sc- 133555783_1	104	104	Unknown, 321 bp.		Novel
NU	scr_sc- 133678871_1	105	105	Unknown, 92 bp.		Novel
NU	scr_sc- 133725675_1	106	106	Unknown, 94 bp.		Novel
NU	scr_sc- 133955481_1	107	107	Unknown, 343 bp.		Novel
NU	scr_sc- 134521597_1	108	108	Unknown, 238 bp.		Novel
NU	scr_sc- 172126480_1	109	109	Unknown, 247 bp.		Novel
NU	scr_sc- 172130231_1	110	110	Unknown, 196 bp.		Novel
NU	scr_sc- 172755010_1	111	111	Unknown, 457 bp.		Novel
NU	scr_sc- 188295137_1	112	112	Unknown, 85 bp.		Novel
NU	scr_sc- 190079504_1	113	113	Unknown, 241 bp.		Novel

1					
NU scr_sc-191455923_1	114	114	Unknown, 388 bp.		Novel
NU scr_sc-195460151_1	115	115	Unknown, 444 bp.		Novel
NU scr_sc-198205946_1	116	116	Unknown, 135 bp.		Novel
NU scr_sc-2573087_1	117	117	Unknown, 246 bp.		Novel
NU scr_sc-2585074_1	118	118	Unknown, 203 bp.		Novel
NU scr_sc-8571871_2	119	119	Unknown, 233 bp.		Novel
NU scr_sc-87731837_1	120	120	Unknown, 300 bp.		Novel
NU scr_sc-87869413_1	121	121	Unknown, 351 bp.		Novel
NU scr_gb-ai233262_2	122	122	Unknown, 889 bp.		Novel
NU cgrrs0h0310.9_13952-135	123	123	Unknown, 310 bp.		Novel
NU scr_gb-m13100.5_2	124	124	Unknown, 100 bp.		Novel
NU scr_sc-170396977_1	125	125	Unknown, 350 bp.		Novel
NU scr_sc-14059147_2	126	126	Rat gene patent WO0210453, 254 bp.		Unknown
NU scr_sc-87750810_1	127	127	Rat gene patent WO0210453, 1063 bp.		Unknown
NU cszr_202034260_190929676	128	128	Rattus norvegicus Tclone4 [U30788].		Unknown

Using the TOXMARKER gene information listed in Table 4, zone 3 necrosis-related genes expressed in vitro were confirmed. Confirmed genes are listed in Table 5.

5 **Table 5**

Gene ID	TOX Number	SEQ ID NO	Definition	Human Ortholog	SEQ ID NO:
<b>Amino Acid Metabolism</b>					
cszr_96561134_83760493	35	39	Rattus norvegicus Carbamoyl-	CPS1: Carbamyl phosphate synthetase I [D90282, NM_001875]	129

		phosphate synthase [ammonia] (CPSASE I), mitochondrial precursor [P07756].	
<b>Carbohydrate Metabolism</b>			
scr_gb- bi277612_1	38	38  Rat gene fragment - 1381 bp. 89% SI (816/910) to Mus musculus for beta- hexosaminidase [Y00964].	NM_000521 >rshd:REFSEQHUMAN DNA- ID:NM_000521 acc:NM_000521 /geneName="HEXB" /definition="Homo sapiens hexosaminidase B (beta polypeptide) (HEXB), mRNA." /protein_id="NP_000512.1" /organism="Homo sapiens" /CDS="76..1746", 1857 bp. >gb:GENBANK-ID:AF378118 acc:AF378118.1 Homo sapiens cervical cancer proto-oncogene 7 mRNA, complete cds - Homo sapiens, 1892 bp.
cszr_229800465_ 190907286	129	131  Rat non-neuronal enolase (NNE) (alpha-alpha enolase, 2- phospho-D- glycerate hydrolase) [X02610].	
<b>Energy Metabolism</b>			
scr_gb-j05266_3	39	.39  Rattus norvegicus mitochondrial H+ ATP synthase alpha subunit [J05266].	NM_004046 >rshd:REFSEQHUMAN DNA- ID:NM_004046 acc:NM_004046 /geneName="ATP5A1" /definition="Homo sapiens ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle (ATP5A1), mRNA." /protein_id="NP_004037.1" /organism="Homo sapiens" /CDS="912..2573", 2725 bp.
<b>Cell Cycle Regulation (Regulation Of Proliferation)</b>			
scr_gb- ab015747_3	43	43  Rat interleukin-4 receptor (membrane- bound form)(AB015747. 1: 99%/3501, p=0.000000), 3520 bp.	X52425.1 Human IL-4-R mRNA for the interleukin 4 receptor
scr_gb- ai233262_2	122	122  Rat gene fragment, 889 bp, 93% identical over 679 bp to mouse RANBP4 [AF123388].	weakly similar to PUTATIVE IMPORTIN BETA-4 SUBUNIT [AK027871, NM_024658].
scr_gb- bi294409_1	42	42  Rat gene fragment, 526 bp, 89% SI (327/365) to mouse type I interferon receptor, IFNaR2	ifnar2: interferon receptor ifnar2-1 [L41944, NM_000874]

			subunit, 1109 bp (Y09813).		
scr_gb-m37394_5	40	40	Rattus norvegicus epidermal growth factor receptor (Egfr) [M37394].	EGFR: epidermal growth factor receptor [K03193, NM_005228].	136
scr_gb-m64300_4	41	41	Rattus norvegicus extracellular signal-related kinase (ERK2) [M64300].	MAPK1: Mitogen-activated protein kinase 1 [BC017832, NM_002745].	137
scr_sc-191879433_1	44	44	Rat Crk-associated substrate, p130 [D29766].	BCAR1: breast cancer anti-estrogen resistance 1 [AK026121, NM_014567].	138
<b>Cellular Communication</b>					
scr_gb-m13100.5_2	124	124	Rattus norvegicus gene for histamine N-methyltransferase, exon 1 and 2 [AB007833].		
scr_gb-x87157_5	46	46	Rattus norvegicus neurotensin endopeptidase [X87157].	NLN: neurolysin [AJ300837, NM_020726].	139
<b>Detoxification Response / Biotransformation-TOX</b>					
scr_gb-af017393_2	48	48	Rattus norvegicus cytochrome P4502F4 (CYP4502F4) [AF017393].	CYP2F1: cytochrome P450, subfamily IIF, polypeptide 1 [J02906, NM_000774].	140
scr_sc-134241980_1	49	49	Rat cytochrome P450IIB3 (P450IIB subfamily) mRNA, complete cds (M20406.1: 100%/279, p=5.0e-156), 367 bp.	MSRA: methionine sulfoxide reductase A [AJ242973, NM_012331].	141
<b>DNA Metabolism</b>					
scr_gb-bi296376_1	94	94	Rat ribosomal DNA external transcribed spacer 1 (ETS1) [X16321], Rat 45S rDNA gene transcription initiation region [X00677], and Rat ribosomal		



			RNA 1.6 small subunit (SS1.6) gene, 3' end [M30843] - 2695 bp.	
<b>Immunity And Defense</b>				
scr_gb-x17037_2	51	51	Rat OX40 antigen [X17037].	TNFRSF4: tumor necrosis factor receptor superfamily, member 4 [AW293499,S76792,X75962,NM_003327]. 142
cgrrs0h0310.9_13 952-135	123	123	Rat lipocortin-V mRNA, complete cds (M21730.1: 99%/1419, p=0.000000), 1744 bp.	
<b>Intracellular Transport</b>				
scr_gb-aj000696_5	53	53	Rattus norvegicus novel kinesin-related protein, KIF1D [AJ000696].	KIF1C: kinesin family member 1C [AB014606, NM_006612]. 143
scr_gb-d79221_3	54	54	Rattus norvegicus r-sly1 [U35364].	Vesicle transport-related protein [AF110646, NM_016163]. 144
<b>Oxidative Stress-TOX</b>				
cszr_229602935_183895355	56	56	Rat metallothionein-i (mt-1)	
m61937	55	55	Rattus norvegicus dihydrodiol dehydrogenase [M61937].	
scr_gb-af069306_1	59	59	Rattus norvegicus transaldolase [AF069306].	clone MGC:2838 IMAGE:2966784 [BC001517]. 145
scr_gb-af106944_3	57	57	Rattus norvegicus Peroxiredoxin III [AF106944].	
scr_gb-d17310_4	60	60	Rattus norvegicus steroid 3-alpha-dehydrogenase [D17310].	MTCH2: mitochondrial carrier homolog 2 [AF085361, NM_014342] 146
scr_gb-m11794_3	58	58	Rattus norvegicus metallothionein-2 and metallothionein-1 genes [M11794].	MT2A: metallothionein 2A [S52379, NM_005953]. 147
<b>Protein Metabolism</b>				

scr_gb-bf281368_2	61	61	Rat gene fragment - 1086 bp. 80% SI (754/938) to Human Prt1 homolog [U62583].	elF3: eukaryotic translation initiation factor 3 [U78525, NM_003751]	148
<b>RNA metabolism</b>					
scr_gb-ai406674_1	63	63	Rat gene fragment, 796 bp. 91% SI to human heterogeneous nuclear ribonucleoprotein C (C1/C2), [XM_166936]	HNRPC: heterogeneous nuclear ribonucleoprotein C [BC003394, NM_004500].	149
scr_gb-bi288503_1	65	65	Rat gene fragment 456 bp. 94% SI (430/456) to Mus musculus cardiac lineage protein 1 (Clp1) (AY090614.1)	HEXIM1 protein [AB021179, NM_006460].	150
scr_gb-d86383_2	66	66	Rattus norvegicus Hex [D86383].	HHEX: hematopoietically expressed homeobox [BC015110, X67235, NM_002729].	151
<b>Membrane Transport / Transporters</b>					
scr_sc-8571871_2	119	119	Rattus norvegicus Na/Pi cotransporter 4 [AB025224.1].	U90545.1 HSU90545 Human sodium phosphate transporter (NPT4) mRNA, complete cds	152
<b>Other</b>					
scr_gb-af311311_2	130	153	Rattus norvegicus P116RIP mRNA, complete cds (AF311311.1: 99%/3285, p=0.000000), 3905 bp.		
scr_gb-ai013477_2	69	69	Rattus norvegicus VL30 element [M91234].		
scr_gb-bm986259_1	97	97	Rat gene fragment, 1047 bp, 86% SI (647/748) to human CGI-126 protein mRNA, 1091 bp (AF151884).	HSPC155: hypothetical protein HSPC155 [AF161504, NM_016406]	154
cszr_230290139_182026368	131	155	Similar to Mus musculus, feminization 1 a		

			homolog (C. elegans), clone MGC:6309 IMAGE:2811079, mRNA, complete cds (BC009161.1: 91%/244, p=3.4e-085), 249 bp.		
scr_gb-m91235_3	70	70	Rattus norvegicus VL30 element [M91235].		
<b>Novel</b>					
scr_gb-aa801331_1	75	75	Unknown, 1252 bp.		
scr_gb-aa899865_3	76	76	Unknown, 1241 bp.		
scr_gb-ai411514_4	79	79	Unknown, 1221 bp.	KIAA0652 gene product [BC006191, NM_014741].	156
scr_gb-bf558463_2	88	88	Rat gene fragment, 921 bp.	Clone IMAGE:4052356, partial cds. [BC014348].	157
scr_gb-bi278552_1	91	91	Unknown, 1060 bp.	KIAA0427 gene product [AB007887, NM_014772].	158
scr_gb-bi295938_1	83	83	Unknown, 1241 bp.	Hypothetical protein FLJ13409 [BC015897, NM_024617].	159
scr_gb-bm384392_1	85	85	Unknown, 2423 bp.	Hypothetical protein MGC3067 [BC002457, NM_024295].	160
scr_gb-s69874_5	98	98	Unknown, 1191 bp.	COL4A1: collagen, type IV, alpha 1 [AH002741, NM_001845].	161
scr_sc-132947646_1	102	102	Rat gene fragment, 50 bp.		
scr_sc-172126480_1	109	109	Unknown, 247 bp.		
scr_sc-188295137_1	112	112	Unknown Gene, 85 bp.		
scr_sc-191455923_1	114	114	Unknown, 388 bp.	Heterogeneous nuclear ribonucleoprotein M [NM_005968].	162
scr_sc-195460151_1	115	115	Unknown, 444 bp.		
<b>Unknown</b>					
cszr_204152648_191521095	71	71	UI-R-A1-ek-e-09-0-UI.s1 UI-R-A1 Rattus norvegicus cDNA clone UI-R-A1-ek-e-09-0-UI 3', 63 bp.	XM_114110 >rshd:REFSEQHUMAN DNA-ID:XM_114110 acc:XM_114110 /geneName="LOC200081" /definition="Homo sapiens similar to muscle-derived protein MDP77 isoform 2 (LOC200081), mRNA." /protein_id="XP_114110.1" /organism="Homo sapiens" /CDS="94..873", 3962 bp.	163
cszr_204152792_191517979	72	72	Rat gene fragment, 131 bp, 99% SI (130/131) to mouse IMAGE	AB032968.1 Homo sapiens mRNA for KIAA1142 protein, partial cds	164

			clone (BC030389), 2072 bp.		
scr_gb- bf390383_1	87	87	UI-R-CA1-bcg-d-03-0-UI.s1 UI-R-CA1 Rattus norvegicus cDNA clone UI-R-CA1-bcg-d-03-0-UI 3', 486 bp. [BF390383.1].		
scr_gb- bf560709_1	89	89	UI-R-C1-kd-h-12-0-UI.r2 UI-R-C1 Rattus norvegicus cDNA clone UI-R-C1-kd-h-12-0-UI 5', mRNA sequence(BF560709.1: 99%/464, p=9.2e-264), 525 bp.	CTNNA2: Catenin (cadherin-associated protein), alpha 2 [M94151, NM_004389].	165
scr_sc- 14059147_2	126	126	Rat gene fragment - 254bp. 75% SI (190/252) to Homo sapiens mRNA for putative progesterone binding protein [AJ002030].	>gb:GENBANK-ID:HSAJ2030 acc:AJ002030.1 Homo sapiens mRNA for putative progesterone binding protein - Homo sapiens, 1874 bp.	166
scr_sc- 172130231_1	110	110	Rat gene fragment - 196 bp. 92% SI (181/196) to Mus musculus cDNA clone IMAGE:5051929 5'[BI146266.1].	AK056165.1  Homo sapiens cDNA FLJ31603 fis, clone NT2RI2002654	167

#### Example 6: Genes Involved in Zone 3 Necrosis *in vivo*

There is ample evidence to support the importance of ATP depletion as a mechanism that leads to cell necrosis [3]. Rat trihydroxycoprostanoyl-CoA oxidase was selected as an *in vivo* marker for zone 3 necrosis and was found to be downregulated in the dataset. This enzyme is one of the three acyl-CoA oxidases found in rat liver peroxisomes and is responsible for the beta-oxidation of fatty acids [16]. Downregulation of an enzyme involved in beta-oxidation represents an interruption in a pathway that can lead to ATP production. Consistent with this is the downregulation of an enzyme involved in amino acid metabolism. Betaine-homocysteine methyltransferase (BHMT) catalyzes the transfer of an N-methyl group from betaine to homocysteine to form dimethylglycine and methionine, respectively [14]. Interestingly, alterations in methionine metabolism have been known to be associated with human liver

cirrhosis for many years [15]. As discussed earlier, severe necrosis is involved in the pathogenesis of cirrhosis [9]. Thus, necrosis itself can lead to an inflammation mediated injury. A marker involved in an inflammation and defense response was found to be upregulated in this *in vivo* marker set. Rat MHC-associated invariant chain gamma stabilizes MHC class II molecules that are at the surface of antigen presenting cells [26]. Rat MHC-associated invariant chain gamma is predicted to be a secreted protein based upon PSORT, SIGNALP, and Hydropathic Profile (HP) analysis

Rat ephrin type-B receptor 1 precursor (EphB1) was downregulated as compared to controls. Upon binding to its ligand, EphB1 has been shown to activate c-Jun kinase through recruitment of an intermediate protein Nck [18]. c-Jun is an oncogene involved in cell proliferation. Regulation of proliferation is an integral part of necrotic cell death, whether it results from compensatory liver regeneration of healthy cells or cell cycle arrest of unhealthy cells. Indeed, ribosomal protein S6 was selected as a marker and was found to be upregulated. It has been known for 20 years that the ribosomal protein S6 is quickly phosphorylated when cells are stimulated to grow or divide [19]. Phosphorylation of S6 occurs in response to mitogens by activation of one or more protein kinase cascades, including MAP kinases [20]. Rat annexin II belongs to a family of proteins that in the presence of  $\text{Ca}^{2+}$  bind to acidic phospholipids. They may also cross-link plasma membrane phospholipids with actin and the cytoskeleton, and possibly play a part in exocytosis, since they are also involved in granule aggregation and membrane fusion [28]. Annexin II was found to be an upregulated marker and may be indicative of loss of structural integrity within the cell. Though annexin I and annexin II have also been identified as major substrates for the tyrosine kinase activity associated with epidermal growth factor receptor (Egfr). Rat Egfr signals through a number of pathways, including the MAP kinase pathway, to regulate proliferation [35].

Canalicular multispecific organic anion transporter (cMOAT), an upregulated gene in this *in vivo* marker set, has been shown to arbitrate the hepatobiliary elimination of many organic ions [29]. It has also been shown to remove glutathione conjugates from cells [30]. Glutathione conjugation is a cellular adaptation to the generation of reactive oxygen species (ROS) [36]. ROS occurs when oxidative phosphorylation becomes uncoupled during ATP generation. When molecular oxygen is reduced in such a manner damaging amounts of  $\text{O}_2^-$ ,  $\text{H}_2\text{O}_2$  and  $\text{OH}$  are

formed in the process. ROS are thought to participate in necrosis through their reaction with all forms of biological macromolecules including lipids, proteins, nucleic acids and carbohydrates [37-39].

Markers were chosen from other biochemical pathways as well. This includes the upregulation of rat heat shock protein 86 (hsp86) which may be indicative of cellular stress. Rat ras-related protein (Rab-2), a protein involved in intracellular transport, was also upregulated. Two protease inhibitors, rat homologue to human inter-alpha-trypsin inhibitor heavy chain (ITI1) and rat homologue to mouse serine proteinase inhibitor mBM2A were downregulated and upregulated, respectively. Human ITI is found in human serum and is predicted to be a secreted protein based upon PSORT, SIGNALP, and HP analysis [32]. Finally, numerous genes involved in xenobiotic metabolism were diminished after exposure to the zone 3 necrotic agents. However, two, rat epoxide hydrolase and rat aflatoxin B1 aldehyde reductase, were found to be upregulated.

There were 11 markers chosen for this PTS marker set that did not match any known genes in the database and have novel composition. One gene, rat osteoactivin has no known association to any of the histopathologically relevant biochemical or toxicological pathways but is predicted to be a secreted protein based upon PSORT, SIGNALP, and HP analysis.

#### **Example 7: Genes Involved in Zone 3 Necrosis *in vitro***

Several of the biochemical events consistent with necrosis are represented in the *in vitro* marker set that is predictive of zone 3 necrosis. There is ample evidence to support the importance of ATP depletion as a mechanism that leads to cell necrosis [3]. Included in this is documentation that the production of ATP via glycolysis can protect a cell from necrosis when oxidative phosphorylation is inhibited [51-53]. The alpha subunit of rat ATP synthase was found to be downregulated in the *in vitro* zone 3 necrosis marker set. ATP synthase is the final enzyme in the electron transport chain and is ultimately responsible for catalyzing the synthesis of ATP. Downregulation of such a key enzyme is indicative of loss of ATP within the cell. Rat nonneuronal enolase is another marker that was found to be downregulated. Enolase is a vital enzyme in the glycolysis pathway that converts glucose to pyruvate. Glucose is a preferred carbon source and generated the highest return of ATP per unit of expended energy. Downregulation of enolase may represent a depletion of glucose stores within the cell. The rat

homologue to mouse beta-hexosaminidase is a protein involved in oligosaccharide and glycosaminoglycan degradation and was found to be upregulated in this marker set [20].

Upregulation of this marker may represent the cells attempt to maintain glucose supplies. The rat homologue of this gene was found to be a secreted protein based on protein based upon PSORT,

5 SIGNALP, and HP analysis. Two markers related to ATP depletion were found to be upregulated in this marker set. Rat EP3 alpha receptors for prostaglandin has been found to be involved in the inhibition of adenylyl cyclase, which catalyzes the conversion of ATP to cAMP [18]. Inhibition of this process would be consistent with a lack of ATP within the cell.

Carbamoylphosphate synthase is a mitochondrial protein that removes excess ammonia in the  
10 cell via the urea acid cycle. Upregulation of this rat enzyme may indicate a resort to utilize amino acids as a source of energy.

Uncoupling of electron transport during oxidative phosphorylation in the above process can lead to the formation of excessive amounts of Reactive Oxygen Species (ROS). When molecular oxygen is reduced in such a manner damaging amounts of  $O_2^-$ ,  $H_2O_2$  and  $OH$  are

15 formed in the process. ROS are thought to participate in necrosis through their reaction with all forms of biological macromolecules including lipids, proteins, nucleic acids and carbohydrates [54-56]. Cells have adapted to the generation ROS through an elaborate antioxidant defense system. Two such mechanisms of defense are found to be upregulated in this marker set. A rat

metallothionein isoform represents one of these mechanisms. Metallothionein is a small  
20 cysteine-rich metal binding protein that mediates heavy metal response and can play a role in ion homeostasis has the ability to scavenge free radicals and has been found to be induced under oxidative stress conditions [39]. Transaldolase is a key enzyme in the nonoxidative branch of the

pentose phosphate pathway that can reduce the amount of reactive oxygen intermediates though the maintenance of glutathione at a reduced state [41]. Glutathione is critical for scavenging

25 mitochondrial ROS through glutathione reductase and peroxidase systems. Interestingly rat peroxiredoxin III, a member of a novel family of anti-oxidative proteins, was found to be downregulated in this marker set. Peroxiredoxins have the ability to reduce  $H_2O_2$  by using thioredoxin or glutathione as an electron donor [40]. Downregulation of this protein may

represent a preference to remove  $H_2O_2$  through one of the other defense mechanisms available to  
30 the cell. Similarly rat metallothionein 1, another isoform of metallothionein, was downregulated.

The fact that there are two rat metallothionein isoforms found in this marker set modulated in opposite directions may also indicate a preference of one form over the other or may represent a redundancy in the pathway. Dihydrodiol dehydrogenase is a marker, which was found downregulated and may play a role in the amount of ROS generated in the cell. This enzyme is believed to yield ROS upon detoxification of polycyclic aromatic hydrocarbons [38]. Thus downregulation of this process may be an attempt to limit the overall amount of ROS within the cell.

As discussed earlier, hepatic regeneration is a response to cellular necrosis. This process involves re-entry of surviving liver cells into the cell cycle to replace lost tissue mass [57].

Though this normal reaction to liver injury can, if uncontrolled, lead to the early onset of hepatic carcinogenesis. Several markers predictive of *in vitro* zone 3 necrosis were found to be involved in the regulation of cellular proliferation in the cell. These markers were consistently upregulated and are represented by growth factor receptors (rat Egfr), cytokine receptors (rat IL-4r & a rat gene homologue to Mus musculus Inar-2 receptor), MAPK signaling cascades (rat Erk2), as well as a gene involved in the regulation of protein kinase C activity (rat homologue to human DGK-delta). Rat epidermal growth factor receptor signals through a number of pathways, including the MAP kinase pathway, to regulate proliferation. However, under certain conditions stimulation of this pathway can lead to cell growth arrest and the induction of apoptosis [22]. Interestingly rat extracellular signal-related kinase 2 (Erk2) was also found to be upregulated. Erk2 is a member of the Raf/MEK/ERK signaling pathway that was the first MAP kinase cascade to be characterized [23]. Rat interleukin-4 receptor (IL-4r) is an upregulated marker for *in vitro* zone 3 necrosis. While IL-4 is a cytokine that has immunomodulatory effects, there is evidence that IL-4 interaction with its receptor can lead to such downstream effects as gene activation and cellular proliferation [26]. A rat homologue to mouse soluble isoform precursor type I interferon receptor (Ifnar-2) represents the upregulation of another cytokine receptor. Recent data shows that murine Ifnar-2 is an effective regulator of interferon responses [24]. It is known that type I interferons play a role in cell proliferation [25]. Murine Ifnar-2 soluble form is predicted to be a secreted protein based upon PSORT, SIGNALP, and Hydropathic Profile (HP) analysis. A rat homologue to human diacylglycerol kinase, delta (DGK-delta) was found to be upregulated in this marker set. DGK is a signal transduction



enzyme that mediates protein kinase C activity by modulating intracellular concentrations of two signaling lipids, diacylglycerol and phosphatidic acid [29]. Protein kinase C is a family of serine-threonine kinases that is known to regulate proliferation and apoptosis [30]. The only downregulated marker involved in regulation of cell proliferation was rat Crk-associated substrate (Cas) p130, a unique docking protein with a Src homology 3 (SH3) domain. Tyrosine phosphorylation of Cas has been implicated in integrin mediated activities including cell proliferation and survival [27]. Additionally, Cas, upon interaction with Src, has been shown to be involved in a H<sub>2</sub>O<sub>2</sub> activation of cJun NH(2) terminal kinase (Jnk) pathway [28].

Some markers may have a less clear association with a necrosis specific pathway. This includes the upregulation of rat VL30 element, a retrotransposon that has been found to be upregulated in rodent malignancies but which a specific role has not been identified [50]. Rat OX40 antigen was found to be downregulated. OX40 is a member of the tumor necrosis factor family that is expressed by activated T lymphocytes and may indicate the presence of inflammatory events [34]. Also relevant to an immunological response is the upregulation rat homologue to Homo sapiens IgG Fc binding protein. This protein is widely expressed on mucosal surfaces and in external secretions [35]. The rat homologue to human IgG Fc binding protein is predicted to be a secreted protein based upon PSORT, SIGNALP, and HP analysis. This marker set includes two markers involved with protein metabolism. This includes the upregulation of rat ubiquitin-conjugating enzyme (UBC), which catalyzes the covalent attachment of ubiquitin to a target protein. The ubiquitin/proteasome pathway is the main non-lysosomal route for intracellular protein degradation in eukaryotes. It is important to many cell processes including cell-cycle progression and more recently has been found to target regulatory molecules found in the apoptotic cell death pathway [44]. Another protein that is involved with protein metabolism and was found to be downregulated is a rat homologue to human Prt1. Little is known of this protein, except that it is a component of the translation initiation factor eIF-3 [43]. Rat thymosin beta-10 is a marker that was downregulated. It is a protein that is believed to be involved in the sequestering of actin and may be indicative of loss of structural integrity of the cell [48].

Markers were chosen from several other biochemical pathways as well. A couple of genes involved in xenobiotic metabolism were induced after exposure to the zone 3 necrotic

agents. These are rat cytochrome P-450's 2F4 and 2B3, which were both upregulated. Several genes that are involved with the synthesis or transport of RNA were also chosen as markers. These include the upregulation of the rat homologue to mouse heterogeneous ribonucleoprotein C, rat homologue to mouse cardiac lineage protein 1 and rat Hex. The rat homologue to mouse heterogeneous ribonucleoprotein G was also chosen as a marker but was found to be downregulated. Two markers involved in the intracellular trafficking were also selected. This includes the upregulation of a rat novel related kinesin protein which may be involved in the transport of membrane bound organelles and the downregulation of rat r-sly1 which plays a role in ER to Golgi trafficking [36, 37]. Two markers in this set are involved with cellular communication. Rat neurotensin endopeptidase was downregulated. Neurotensin is a hormonal peptide that functions as a central neurotransmitter or neuromodulator as well as a peripheral hormone [20]. Rat densin-180 has a strong association with the postsynaptic density in CNS neurons and is believed to function as a synaptic adhesion molecule. Densin-180 was found to be upregulated in this *in vitro* marker set [31]. Rat homologue to mouse sex-determination protein homologue Fem1a was upregulated in this marker set but has no known association to any of the histopathologically relevant biochemical or toxicological pathways. Fem1 is a known signal transducing regulator in the *C. elegans* sex-determination pathway [49].

There were 55 markers chosen for this PTS marker set that did not match any known genes in the database and have novel composition. There were also two genes that blasted to rat gene patents and have novel utility. One additional marker, rat Tclone4, had some similarity to a known gene but has no known function.

### **EXAMPLE 8: Prediction of the Toxicity of a Test Compound**

The following example describes the application of the TOXMARKER expression profiles generated, as described above, to identify hepatotoxic compounds.

#### **Hepatocyte Culture**

**Animals.** Male Wistar Han rats (CrI:WI[Glx/BRL/Han]IGS BR) were obtained from Charles River Laboratories, Inc. (Raleigh, NC). The animals were housed for 6 or 7 days in a

temperature-, humidity-, light-controlled facility prior to hepatocyte preparation and were at this time 200-250 g in weight.

**Hepatocyte isolation** Hepatocytes were prepared by in situ liver perfusion according to a protocol used at the Yale Liver Center (Yale University School of Medicine, Yale University, New Haven, CT). To minimize the risk of contamination all equipment and solutions used during the perfusion procedure were autoclaved. The animals were anesthetized by sodium phenobarbital (approx. 50 mg/kg) and the abdomen opened to expose the liver. To perform the perfusion a catheter was attached to vena porta and secured by a ligature. After disrupting vena cava inferior 37 °C Hanks A buffer (120 mM NaCl, 5 mM KCl, 0.4 mM KH<sub>2</sub>PO<sub>4</sub>, 25 mM NaHCO<sub>3</sub>, 0.5 mM EGTA, 0.1% glucose) was circulated through the liver a 40 ml/min for 10 min. The perfusion was continued with 37 °C Hanks B buffer (120 mM NaCl, 5 mM KCl, 0.4 mM KH<sub>2</sub>PO<sub>4</sub>, 25 mM NaHCO<sub>3</sub>, 0.4 mM MgSO<sub>4</sub>, 0.5 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>, 0.1% glucose) supplemented with collagenase (Liberase Blendzyme 3, 120mg/400 ml; Roche Diagnostics Corp., Indianapolis, IN) until clear signs of liver disintegration were visible (approx. 10 min). The liver was transferred into ice-cold Leibovitz L-15 media (Invitrogen, Carlsbad, CA) and the organ was disrupted mechanically with forceps. The cell suspension was filtered through a 80 µm mesh followed by a 45 µm mesh, and the medium replaced twice with fresh ice-cold L-15 medium following 5 min low speed centrifugations (30G). Cells were transported on ice to Curagen's tissue culture facility and washed twice as above in William's E media (Sigma, St. Louis, MO) with supplements (10% fetal bovine serum (Gemini, Woodland, CA); 9.6 µg/ml prednisolon, glucagon 0.014 µg/ml, insulin 0.16 units/ml, glutamin and antibiotic-antimycotic solution (all purchased from Sigma, St. Louis, MO). Cells were counted and viability was measured in a hemacytometer after Trypan blue staining

**Hepatocyte culturing** Cells were cultured in 12 or 24 well tissue culture plates coated with rat tail collagen (Becton-Dickinson, Bedford, MA) according to a protocol provided by Dr. Grazyna Wasinska-Kempka at Bayer (Wuppertal, Germany). Cells were seeded at a density of approximately 80,000 cells/cm<sup>2</sup> or 270,000 cells/well in 12 well plates and 140,000 cells/well in 24 well plates and incubated at 37 °C, 5% CO<sub>2</sub> for 2 h. To suppress contamination the amount of antibiotic-mycotic solution was increased (3-fold) and gentamicin (44 µg/ml; Invitrogen, Carlsbad,) was added during the seeding stage. After 2 h the media was replaced with fresh

William's E media (as above) supplemented with 75 ug/ml rat tail collagen (Becton-Dickinson). The media was replaced approximately 16h later with fresh collagen-containing media, with or without test compound, and thereafter every 24 h

**Test compounds** Test compounds were solubilized at 200-fold final concentration in 100% DMSO and diluted into hepatocyte culture media to final concentrations ranging from approximately 200 pM to 10 mM depending on the compound. Compound stock solutions in DMSO were prepared at the initiation of the dosing regimen and stored at 4 °C. Compounds were added to the culture media immediately prior to addition to the cells.

Control compounds (compounds that do not elicit a toxic histopathology, see for example Table 1 and 2)) are selected from the training set and dosed on each batch of hepatocytes along with the unknown compounds subjected to toxicity prediction. The data from these controls are used to determine hepatocyte quality over time

**RNA isolation** Qiagen's (Valencia, CA) RNeasy 96 kit was used for isolation of RNA. For cell lysis the culture media was carefully removed and 400 ul/sample in 12 well plates and 200 ul/sample in 24 well plates of RLT lysis buffer supplemented with 10 mM DTT was added per well. The RLT buffer efficiently lysed cells and solubilized the collagen layer covering the cells. The lysates were homogenized by pipetting 12-15 times, snap frozen and stored at -80 °C.

**Cell viability** In parallel with the compound dosing, a viability assay was performed to make sure that the concentrations used to treat the hepatocytes were not excessively toxic to the cells. Cell viability was monitored for each compound concentration using the CellTiter Assay (Promega, Madison, WI), a modified MTT assay. Cells were seeded in 96 well plates coated with rat-tail collagen (Becton-Dickinson) at a density of 27,000 cells/well. After 16 h in culture cells were treated in triplicate for each compound and concentration. Following 24, 48 or 72 h incubation in the presence of the drug, the MTS/PMS reagent of the CellTiter kit was diluted 1/6 in culture medium, added to the cells and after 60 min incubation of the cells optical density at 490 nm was measured using a PowerWaveX Select 96 well spectrophotometer (Bio-Tek Instruments, Winooski, VT). Viability was calculated in relation to no-drug control after

subtraction of a no-cell background value. Only compound concentrations that showed 70% or more viability in this assay were used for gene expression profiling.

**RNA Purification** The hepatocytes are harvested in 200 µl of lysis buffer (RLT) provided with the Qiagen RNA isolation kit. Total RNA is isolated from the lysates using the  
5 Qiagen RNeasy 96<sup>®</sup> isolation kit following the manufacture's instructions with some modification. An equal volume of 70% ethanol is added to each of the lysates and the samples are added to the membrane on the 96 well plate. Membranes are then washed thoroughly (once with 800 µl of buffer RW1 and twice with 800 µl buffer RPE) to remove unbound material followed by DNase I treatment (50 units of DNase I from Promega (10 u/µl) in buffer RDD from  
10 Qiagen; total volume 60 ul) for 1 hr at room temperature to remove all traces of genomic DNA that might be co-purifying with the RNA. Following DNase I treatment, the membranes are again washed three times as before (once with 800 µl of buffer RW1 and twice with 800 µl buffer RPE), and dried with a centrifugation step (6000xg for 7 min; to remove all residual traces of ethanol from the washing buffers). RNA is then subsequently eluted from the columns with 40 µl  
15 of RNase/DNase-free water. This process has been automated using the Tecan Genesis Freedom robotic system.

Quantity of RNA is determined by fluorometry using Ribogreen dye from Molecular Probes and quantified using a fluorometer (Spectrafluor Plus instrument, Tecan). This procedure involves diluting each 2µl sample ten-fold in Rnase-free water and then measuring  
20 fluorescence (after addition of dye). An average of triplicate subsamples is used to calculate the concentration and total RNA yield for each sample (by comparison to a standard curve generated from known amount of RNA standards). At this point the samples are evaluated as passed or failed based on a concentration criteria (60 ng/µl or more considered as "PASS"). The samples that have a concentration greater than 60 ng/µl are further diluted to 60 ng/µl with DEPC treated  
25 water.

**cDNA Synthesis** Double stranded cDNA is synthesized using the Roche cDNA synthesis kit, following the manufacturer's instructions, with some modifications. 600 ng of total RNA isolated from the hepatocytes (60 ng/ul), are spiked with 2 µl of reference mRNA ( $7 \times 10^6$  copies of *hyaB* and  $2.5 \times 10^7$  copies of *mhpR*) and this mix is incubated in the presence of 2 ug of

oligo [(dT)<sub>24</sub> T7prom]<sub>65</sub> primer at 70°C for 10 min, immediately followed by quick chilling on ice. To each sample first strand synthesis mix is added such that the final mix contains 1X AMV RT buffer, 8.1 mM DTT, 25 units of AMV reverse transcriptase, 12.5 units of RNase-inhibitor and a dNTP-mix (1 mM of each nucleotide). This mix is incubated at 42°C for 1 hour followed by chilling on ice. The second strand synthesis involves the addition of the second strand buffer to a final concentration of 1 X, a dNTP mix (80 µM each) and the second strand enzyme blend provided in the Roche cDNA synthesis kit. The mix is incubated at 16°C for 2 hours. Adding 10 units of T4 DNA polymerase to each reaction and incubating at 16 °C for a further 5 minutes to terminate the elongation. The reactions are stopped by adding 11 µl of 0.2 M EDTA pH 8.0. The double stranded cDNA is purified using Qiagen's QIAquick™ Multiwell PCR Purification kit, following the manufacturer's instructions. The cDNA is then quantified by fluorometry using the Picogreen® dsDNA Quantification Kit (Molecular Probes) following manufacturers instructions.

#### Microarrays

***In Vitro Transcription, cRNA Purification, and cRNA Quantitation*** The complete yield of double stranded cDNA (minimum 50 ng) is placed in a Centrivap Concentrator (Labconco) for 2 hours at 45°C or until liquid is completely evaporated. In vitro transcription is performed using Ambion's MEGAscript™ T7 Kit, following the manufacturer's instructions, with the following modifications. Biotin labels are incorporated during cRNA synthesis by adding biotin 14-CTP (Invitrogen) and biotin 16-UTP (Roche Applied Science) to the in vitro transcription reaction to a final concentration of 1.5 mM. Transcription reactions are incubated at 37°C during 16 to 18 hours. Reactions are stopped by adding 2 U of DNase 1, and incubating at 37°C for 15 minutes. cRNA is purified using Qiagen's RNeasy 96™ kit, following the manufacturer's instructions. Purified labeled cRNA is recovered in 30 ul of DEPC-treated water (see STM TS-MAH-104) . Labeled cRNA yield and quality are determined by measuring the 260/280 nm optical density ratio, using a Powerwave HT spectrophotometer (Bio-Tek). Labeled cRNA is then diluted to 0.225 ug/ul using DEPC-treated water.

**Hybridization, and Fragmentation** We use a format with two microarrays on each slide and dual hybridization chambers consisting of two 0.8 mm height, 22 mm<sup>2</sup> square chambers. Dual hybridization chambers are installed on glass arrays using a Slide-Chamber Alignment Tool (SCAT). The SCAT is linked to a vacuum pump to create pressure that makes the chamber and the microarray hold together. Once assembled, microarrays are incubated at 37°C for 10 minutes. Slides are then placed chamber side down on a lint free paper towel (Texwipe Company LLC) on a flat surface, and using a finger, gentle pressure is applied along the periphery of the chamber.

Synthetic cRNAs (1 ng each of bacterial araA and ybiW) are combined with 9 ug hepatocyte-derived labeled cRNA prior to fragmentation. The cRNA mixtures are fragmented at 94°C for 20 minutes in a 50 ul reaction containing 40 mM Tris-Acetate, 100 mM Potassium-Acetate, 31.5 mM Magnesium-Acetate, pH 8.1. After the fragmentation is complete, 300 ul of cold hybridization buffer (Mergen Ltd., proprietary composition) is added to the fragmented RNA, and kept on ice until ready for loading onto a microarray. 320 ul of cRNA are then loaded onto a barcoded microarray. After air bubbles have been removed from the liquid, loading ports of the chamber are tightly sealed using Mergen adhesive dots. Slides are incubated for 16-18 hours at 30°C in a hybridization rotisserie oven set at 18 rpm (Robbins Scientific, model 400)

**Microarray Washing and Staining** After hybridization for 16 hours, the hybridization chambers are removed from the microarrays slowly using forceps. The microarrays are placed in a reservoir containing TNT (0.1 M Tris-HCl, pH 7.6, 0.15 M NaCl; 0.05% Tween-20) and incubated at 39°C for 1 hour. The following staining procedure is light sensitive so all incubations are done under foil taking care to minimize light exposure. Following the TNT wash, the arrays are incubated in Buffer B (Mergen LTD, proprietary composition) at 4°C with gentle agitation for 30 minutes. The slides are then placed in a 1:500 dilution of 1 mg/ml streptavidin-alexa 647 (Molecular Probes) in Buffer B at 4°C with gentle shaking for 30 minutes. Following a TNT wash of three times 5 minutes each, the arrays are incubated in a 1:500 dilution of 0.5 mg/ml biotinylated anti-streptavidin antibody (Vector Labs) in Buffer B at 4°C with gentle agitation for 30 minutes. The microarrays are washed three times in TNT for 5 minutes each. A 1:1 mix of streptavidin-alexa 647 and biotinylated anti-streptavidin antibody is incubated at 25°C for 1 hour to encourage complex formation. This complex is diluted 1:267 in Buffer B and

incubated with the microarrays for 30 minutes at 4°C with gentle shaking. The slides are washed again in TNT three times for 5 minutes each followed by 2 washes of 5 seconds each in redistilled H<sub>2</sub>O. The arrays are washed individually in redistilled H<sub>2</sub>O five times for 1 second each. The microarrays are placed in a slide holder and centrifuged for 15 minutes at 300 rpm to ensure complete dryness. The slides are placed in a clean dry box and stored at room temperature until being scanned.

**Scanning and Analysis of Microarrays** Each slide is individually scanned using a GenePix 4000B scanner (Axon Instruments Inc.), using the 630 nm laser at 100% power and a PMT setting of 600 volts.

**Image QC and Data Export** Tiff images from each slide are analyzed using the GenePix 4.0 software (Axon Instruments Inc.). The scanned image is first aligned with a grid consisting of an array of circular *features*, such that each spot on the image is contained within a *feature*. After the array is roughly aligned, the software conducts fine alignment of each feature with each spot on the array. Additionally the software calculates local median foreground and local median background for each feature. The microarray is then evaluated manually to determine whether the features found by the software were true data points and not false intensity due to a defect or contamination. If contamination affects more than 30% of a feature, the feature is flagged as bad. If contamination affects more than 30% of the local background area of a feature and the local foreground/local background intensity ratio is less than 3, the feature is flagged as bad. After flagging is complete, a *GenePix output file* is generated. The output file contains spot IDs, spot location information, median local foreground, median local background, and bad spot flag information.

A quality control log is kept for each array to document background levels, noise/contamination problems as well as the number of features flagged. A final *Image QC* pass or fail decision is made for each array and added to the log. An array must contain no more than 1% flagged spots in order to pass *Image QC*. If an array contains an area of contamination of >150 counts which covers more than 5% of the array then the array fails

**PTS Microarray Data Processing** The *GenePix output files* generated during the *Image QC and Data Export* are processed using Microsoft excel. A macro is used to split the data from



each of the two arrays, calculate spot intensities, and filter out low quality data. The following procedure is used to calculate gene intensity and filter the data.

#### A. Calculate spot intensity

5

$\text{spot\_intensity} = \text{median local foreground} - \text{median local background}$

#### B. Calculate threshold

10

There are 15 probes on the array that are either yeast or mouse negative controls.

If (spot\_intensity is from probe type “yeast\_neg” or spot\_intensity is from probe\_type “mouse\_neg”)

then spot\_intensity = thres\_value

15

$\text{mean\_thres} = \text{average}(\text{thres\_value}[1], \text{thres\_value}[n])$

$\text{stdev\_thres} = \text{stdev}(\text{thres\_value}[1], \text{thres\_value}[n])$

20

If ( $\text{thres\_value} < (\text{mean\_thres} + 3 * \text{stdev\_thres})$  and  $\text{thres\_value} > (\text{mean\_thres} - 3 * \text{stdev\_thres})$ )

then thres\_value = good\_thres\_value

else thres\_value = bad\_thres\_value

$\text{mean\_good\_thres} = \text{average}(\text{good\_thres\_value}[1], \text{good\_thres\_value}[n])$

```

        if mean_good_thres > 0
            then threshold = mean_good_thres + 5* stdev (local_bg_median[1],
local_bg_median[n])
5         else threshold = 5* stdev (local_bg_median[1], local_bg_median[n])

```

### C. Data filtration

If (spot\_intensity has no problem spot flag then spot\_intensity = good\_spot\_intensity

10

Only data of type good\_spot\_intensity is passed to the Discovery department.

***PTS Microarray Quantitative QC*** The array quality is assessed by calculating a set of descriptive statistics and testing if they pass set criteria. In order for array data to be accepted the following criteria must be met:

15

- Toxicity marker spot intensity trim mean / threshold > 8
- Maximum spot intensity of blank spots < 300
- Number of marker spots above threshold > 800
- Labeling spike mean spot intensity/ threshold > 50
- 20 • Hybridization spike mean spot intensity >50

### Toxicity Prediction

The PTS presents a single overall likelihood of toxicity for unknown samples (i.e. the probability that a sample is toxic) that can be ranked to indicate severity of toxic insult. In order to compute the toxicity likelihood of an unknown we calculate a likelihood estimate from 3 different modeling types (Classification Trees, Discriminant Analysis, and Logistic Regression) using three independent gene lists for a total of 9 models per mode. These 9 models are then averaged to provide a likelihood (probability) of toxicity for each mode. The rationale behind combining the results of several models as opposed to relying on a single model that performs best for a given mode is to control for the risk of over-fit (a model that performs well on training data but inadequately on novel samples). The best performing model would be expected to be more over-fit than a weaker performer. In order to reconcile this we average the results of several models in order to determine the true likelihood of toxicity for an unknown. This section attempts to describe the methods we will use for model averaging, explain how compounds can be ranked, and how we can draw inferences about the severity of toxic insult.

#### Common Terms:

*Likelihood of Toxicity:* A value between 0 and 1 indicating how confident we are that a given compound/dose combination is toxic.

*Model:* A statistical algorithm for prediction. This section focuses on Logistic Regression, Discriminant Analysis, and classification trees which are explained elsewhere (Appendix H).

*Model Type:* Logistic Regression, Discriminant Analysis, or classification trees

*Mode:* A specific type of hepatotoxicity (e.g. hypertrophy)

*Present and Absent:* Present means a compound produced a given toxicity *in vivo* absent means it did not.

#### Model Background

One obvious problem with model averaging is that different models provide different outputs. Discriminant Analysis produces a number without bounds, logistic regression

produce a likelihood estimate with a value of 0 to 1, and classification trees a proportion of node impurity with a value between 0 and 1. Therefore an arithmetic mean of these three results may be misleading because the scale of discriminant analysis is so different than the other models.

Our approach is to scale the results of all models to produce a single likelihood,  $P^P$ , the

- 5 probability that this sample belongs to the toxicity class as opposed to  $P^A$ , the probability that the sample does not manifest itself as toxic for this mode. By definition:

$$\text{Equation 1: } P_j^P + P_j^A = 1$$

Where  $j$  is the  $j^{\text{th}}$  model (the three models mentioned above). In order to proceed we first need to find  $P^P$  for each model.

10

Equation 1 means that our classifications (present and absent) are mutually exclusive and collectively exhaustive. If a sample is absent for hypertrophy, it cannot also be present for hypertrophy (mutually exclusive) and if a sample is not hypertrophic it must be absent of hypertrophy (collectively exhaustive).

15

*Logistic Regression:* This model returns  $P^P$  by design.

*Discriminant Analysis:* This model returns a linear discriminant that is a one-dimensional linear combination that establishes two separated normal distributions as follows:

20

where ‘Absent’ refers to the theoretical distribution of samples that did NOT produce pathology and ‘Present’ refers to the theoretical distribution of samples that did produce pathology. Let  $M_P$  = the mean of the linear discriminant function for the training samples annotated as present, and  $M_A$  = the mean of the linear discriminant function for the training samples annotated as absent. Let  $G_P$  and  $G_A$  denote the probability mass function for the Gaussian distributions of the linear discriminant values for the training samples annotated as present and absent, respectively. Having estimated these functions, calculations of percentiles is

25

very straightforward. In the above figure,  $M_P > M_A$ , therefore, for an unknown sample with a linear discriminant value of 'x', we have:

$$\text{Equation 2: } P^P = \frac{\text{LowerTail}(\mathbf{G}_P(x))}{[\text{UpperTail}(\mathbf{G}_A(x)) + \text{LowerTail}(\mathbf{G}_P(x))]}$$

5

where Lower Tail  $\mathbf{G}_P(x)$  refers to the area under  $\mathbf{G}_P(y)$  for which  $y < x$ , and Upper Tail  $(\mathbf{G}_A(x))$  refers to the area under  $\mathbf{G}_A(z)$  where  $z > x$ . When  $M_A > M_P$  the above equation becomes:

$$\text{Equation 3: } P^P = \frac{\text{UpperTail}(\mathbf{G}_P(x))}{[\text{LowerTail}(\mathbf{G}_A(x)) + \text{UpperTail}(\mathbf{G}_P(x))]}$$

10

Equations 2 and 3 appear more complicated than they really are. They simply convert the percentile of (x) belonging to the “present” distribution to a conditional probability that it belongs to present and not absent. This is used to satisfy mutual exclusivity rule of equation 1.

15

*Classification Trees.* As mentioned above, classification trees return a probability of correct classification for each prediction. However, this probability is actually a proportion based on the node impurity of the classification tree (the fraction of training samples on that leaf that belong to the majority class, e.g. a leaf contains 9 samples with hypertrophy and 1 sample without, the probability returned is 0.9). This is not a good indication of the true probability of an unknown because the leaf may have very few members and because the confidences of branch decisions are not included in this calculation. In order to convert this proportion to likelihood, we simply construct many trees using a subset (n-1, where n= the number of compounds) of the compounds in the reference database until all compounds are NOT used once (this is identical to

25

the leave one out cross validation described in Appendix H and in the next section). This process is called a jack-knife estimate of confidence.

Summary:

This section describes how we convert the results of the three different models to a likelihood estimate that satisfies the mutual exclusivity rule of equation 1. For logistic regression this is the result, for discriminant analysis we rely on the probability mass function of a normal distribution, and for classification trees we create a jack-knife estimate of node impurity. Having described these techniques the next section explains how we combine the results of the individual models.

*Model Averaging*

Each of the above models makes a decision as to whether a toxicity is “present” or “absent”, and some models are expected to perform better than others. What remains to be explained is how we determine model quality and how we combine the results of individual models.

a) *Model Quality:*

In order to determine how confident a given model is in its decision, we perform jack-knife estimates of each prediction. A jack-knife estimate computes  $n$  different models, with  $n-1$  compounds in the training set (where  $n$  is the number of compounds used for training a particular toxicity mode). Each jack-knife casts a single vote for absent or present. The best models confidently (likelihood is much greater or less than 0.5) make the same decision consistently, while poorer performing models tend to have equivalent “present” and “absent” votes.

b) *Combining Models:*

The number of present and absent votes are tallied across all the jack knife estimates for each unknown compound. Models that are more confident in their decision are

naturally weighted heavier by an ability to cast more votes. The vote totals can then be evaluated using the binomial distribution as follows:

$$L_{Tox} = \sum_{i=0}^P \frac{\binom{N}{i}}{2^N}$$

5

Where  $L_{Tox}$  is the likelihood that the observed vote distribution is greater than 0.5, P is the number of votes for “present” and N is the total number of votes cast.

#### Summary:

10 This section explains how we use three different marker sets for each model and then compute a weighted average based on how consistent the predictions are within a given step. At the completion of this exercise each sample will have three sets of votes: one each for logistic regression, discriminant analysis, and classification trees. The likelihood for each mode is calculated from a binomial distribution, under the null hypothesis that voting is random.

15

#### **EXAMPLE 9: TOXMARKER Nucleic Acid Sequences**

This example provides exemplary TOXMARKER nucleic acid sequences, useful in methods of screening compounds for hepatotoxicity according to the invention.

**Table 9A**

20 >scr\_gb-af038870\_4 (TOXMARKER Assignment: 1; SEQ ID NO: 1)  
 tttttttttttttttttttgaagggttttcaaccggcatgtttttattaatgaaatggaa  
 tggagcagtcagaacagagattacagaattacagaatggatcagttatctgttaagttt  
 tacagggctggtgtgtgtgtgtttctgcctaagggtcctgctcaaaagatcttggaatcca  
 cttgggaagcatcttagatatagatgggtgctgtgtcacttatgatacgggccctgaatg  
 gttctatgtcactcgtggaggtggtgtcctatccccctatctgaaatgagattgacgtcg  
 25 ggtgactttctcttcgctgcagtgactcctgtgcgcctgtaatgcgacaggcacgtagga  
 aatgtgttcaggatttactgtggacttctcctttcttctccttaggtaaaattctaaagc  
 gtagttttgtaactgtgaaatgctatctgtgactccattttgtctaactagcaccaatca  
 caggtgtaagccggcatcaacacaaacgctggtttagagatgccttctccttcgggtgc  
 30 aactgtggcccgacctggaggaattcgccccgaaccgctggcctgtggctactgtgcg  
 gatttgaattttgtttttcgaagagcgctctcagctgctgctcagtggtggcttccttc  
 tgctgcatcagctctgctgcccccttcgtcactcccaagcatccggcttggaatcgaa

ggattgtacggtctgcccgaagctattcgaagattctgccagttattctttcctggccctt  
 gccctgatccagggtttggtgtgcatgtccaaaccacttcccagctgccatgtttttct  
 gaagctggtggttaaaaatccccctttctggggcgagctcctctgcaatggccctgatgtgg  
 tagggctcaaatccgcagcagccgccaatgtacctgacccccaggttgtaggcctctctg  
 5 gcgtatttttgaatatcccatctggtggcaactctgggttccaatccaaaggggaattct  
 gggagatcaataaatccctgtttgccacagtcaggggtgtggtaggccaggggtggctc  
 atcaagtaagccttcagccgagctgcttccagaccctcctcatgagctttattgtctgc  
 aagctggtgctggggtcgaagtggcagttcacaccgacaatggcggcacctgcttttacc  
 aaacgcactgcgcaactctccaggagacacgccatgtagatctccttcaggtccgatgcac  
 10 atggtagccgctataggcttcccgatgttttaaggcctcgactgccacacggcttct  
 tcaacatgttcaaaatactctgcaatgaggaagtccacattcttcttcatgaagacctca  
 agctgttgggtgaaatatcttttaacttccgtctcactcttgcagctgaggtaggaaggt  
 gtctgactcacacctcctgcaaccaatgcatcccccttcgtcagcaacttgccgtgcaatg  
 tcacaagcagcttcatgtaccttctgccagatatcttctctgccacgtagtccctcgg  
 15 tttccagcttgtcctcacttgcatagaagtgaaaggtctgcatgacgttcgatccagct  
 ctgaggaactcccgatgaagctgccgaactgcctcgggggtgctccaccgagcctctggg  
 gtccagggtccagcctttacgtagccccctctttccagtgcaaagacaaatcccccatct  
 ccgatcacgacttcgccagcatttaagcgttctaagattccccctcttggccttcttgccg  
 gcaatcgggtgccatctttccgggtgtcctgagtggcgctgaacgcagctgcggactggaca  
 20 ggagcggctctccagcaaaggcttgactgctgagccgcttctggcctctttatatacagca  
 gctaggattccccagccttgaccgggtccaacacatggcctcagggcggggaacacgcccc  
 ccagcctttgaaacaggcctggggtagctgggaattc  
  
 >scr\_gb-z83053\_3 (TOXMARKER Assignment: 2; SEQ ID NO: 2)  
 25 gacatggcaccagccggaggccacaggtcaagaagggtatcttgagcgtctggacagc  
 ggggaggttgtggttggggacggcggtttcttctcactctggaagagagaggtttgtg  
 aaggcaggactttggactccagaagcagtggttagagtatccaagtgcagttcgtcagctt  
 cacacagaattcttgagagcgggagccgatgtcttgacagacattcaccttttcggctgct  
 gaagacagaatggaaagcaagtgggaagctgtgaatgcagctgcctgtgacctggccag  
 30 gaggtggctgatggaggggtgctttggtggcagggggcatctgccagacatcactgtac  
 aagtaccacaaggatgaaactagaattaaaaacattttccgactacagctaggtgtttt  
 gccaggaaaaaatgtggacttcttgattgcagagtattttgagcatgtggaagaagccgtg  
 tgggctgtggaagtcttgagagaggtgggggcacctgtggctgtgacctgtgcatcggc  
 ccagagggggacatgcacggcgtgacaccgggagagtgtgagggtgagactgtctcgtgca  
 35 ggggccaacatcattggggtaaaactgccggtttgggcctggaccagcttacaggacctg  
 agctcatgaaggaggccctcagggttgccgctactagctcaccttatgggtccagctgct  
 tgggttttctcacactgggactgtggcaaggagggttgtggacttctgatataccttt  
 cgctggggccaagagtggccaccagatgggatattcaaaaatacggccagagaggcctaca  
 acctgggggtcaggtacattggcggtgctgaggatttgagccctaccacatcaggggccc  
 40 attgcagaggagctcgccccagaaaggggattttggccaccagcttcagaaaaacatggc  
 atctggggaagtgggtttggacatgcacaccaaaccctggatcagagcaagggttagacgg  
 gaatactgggaaactctgttgccagcttcgggaagacctttctgtccttccctatcaaa  
 ccagatgcttgagaagccatgaaagagacctctgaagtgcagaaaggaggaaacagcct  
 caagccccatctggaatcttctggctgctgtcctcagcccgcttcttctggctgttgagc  
 45 atcgatgagctgtcgtcccttccaattgagtgcacatcactcctgagtatgccactag  
 atgcggtggagatgcagaggcatccggaccacgccccacccctccccctcacacactt  
 actctctgcctagtaaatgccacagagcttccatccccatccaaagggtcatcaggcatggc  
 tatcagttggctctcaggggtgatttgacattctcagatgattagaagtggcaagaagc  
 aaccttgggtgaataactctgggtgtctaaactctgtacttgagttacagctctcagtagagg  
 50 agacgcaaaagctgttgcgagtgacggcagaattattgaacagtcagtgcttggcttt  
 caaaggcgattatcgctttaaggctcttagaattagtaagtgcattttataaccaggcat  
 agctagatcataaactactgatggccaaggacctagaacgtgcttcttaccttctctc  
 tagtttagcattacgacaaacataatcaccaacgctcagggaaacacttgctgattcaagt  
 aaaatgcatgaaccttggaagaccttctagaagtgcagatcaagttcatcttggctta  
 55 gcactttccacattcatgtttgggttggatgctgcgccctacttttgggttttttgctacaa  
 tgaacaaaattagtgaataaccattagtgaattgcgaataattttccttttctaaattt  
 tgatttctttggaacattgatttaaaaaaaatagtgtgttgccttgctcaaaaaaaaaaaaa  
 aaaa  
  
 >scr\_gb-x95189\_4 (TOXMARKER Assignment: 3; SEQ ID NO: 3)  
 60 ccatagcgaagacttcatgaagactgtcccaggcatgctgtgacacaaactacagaagggt



5 gggaaaagatcttttgggtcaaaccatccggaccttggtaccgcagacagaacaatact  
 gaccgcattcaactacacacagttctcggcacctcccagtgctcagagcagaccctcaag  
 gagatgagcagatccaggatggggagcccaatgcaccgagtgctccctgggggacacctgg  
 agctggcaagtgcacccggacatagacagcgaaaggcactcacgcctctcagtggtgag  
 10 cgactcaccaacatccttgatggaggcctccaaacacccgtgctgcgaagaaaagtgcga  
 agcatcatacaaagtgaccagtggttaattgaagaagctttacttcatgaccgcagag  
 gagctatatgaggatgcgattcaaaagagattccatctcgagaagctagcctggagcctg  
 ggctggtcagaagatggctcctgaacgcatttatgctaacagagtccttgatggaaacgtc  
 aacttaagcttacatgggtgttgccatgaatgctatccgaagcctgggctcagatgaacag  
 attgctaaatggggccaaactctgcaaaaacttccaaatcatcacacatacgcccagaca  
 gagctgggacacgggacatacctacagggcctggagactgaagccacctatgatgaagcc  
 aggcaggagcttgtgatacacagccctacgatgacttccaccaagtgggtggcctggggac  
 ttgggatggctcgggtcacccatgctgtggtcctagcccagttgacctgcttaggagtcggg  
 15 cacggcatgcacgccttcattgtgcccattcggagcctagaggatcacaccccactgcc  
 ggaatcacagttggggacataggccccaaagtgggtttggaacacatagacaatggcttc  
 ctgcaactgaaccacgtgcgggttcccagagaaaacatgctcagtcgctttgcagaggtc  
 ttgccagatgggtacctaccagaggcttgggacgccacagagcaattatcttggcatgtg  
 gtgacccgggtgcagctgctgtgtaagggaatcctaccctccctccagaaggcttgcac  
 attgccacgcgctactcagtaatccgccatcagtcctcgacttccggccagtgaccagag  
 20 gcaaaaatcctggaataccagacgcagcagcagaaaactccttccctcagcttgctgtgagc  
 tatgccttccacttcacggccaccagcctctcagaattcttccacagctcctacagtgtc  
 attctgaagagagacttcagcctcctgcctgagctccatgcattgagcactggatgaag  
 gccacgtttgcagacttctgtgcccagggcgccgagatctgtcgagagcttgccgggggc  
 catggctactcaaagctgagcggcctgcccagactgggttgcctgagcaacagcctcttgc  
 25 acatatgaggggtgagaatacgggtgctctacctgcaagtggccagggttcttgatgaagagc  
 tatctgcaggctcaagcgtccccaggcgccacaccacagaagcctctccctcagtcgctc  
 atgtatattggcacacaaaaggccagccaggtgctcagcccagactgcagctgacttccgc  
 tggccagatgtctataccacagcctgggcatatgtgtctaccaggctcataagagatgca  
 gcacaccgtacacagaccctcatgaagtccggggttgaccagcatgatgcctggaatcaa  
 30 actactgtcatccaccttcaggctgctaaggctcactgctacttcatcactgtgaagaat  
 ttcaaggaagctgttgagaaaactagacaaggaaccagagattcagcgtgtgctccaacgc  
 ctctgtgacctctatgccttacacggtgttctgactaactcaggggactttctgcatgat  
 ggcttcctgtctggggcccagggtggacatggccagagaagccttccctagacctgcttccc  
 ttgatccggaaggatgccatcttgttaaccgatgcttttgacttctcggaccattgttta  
 35 aactcggcacttggctgttatgatggacacgtctacgaacgcctgtttgagtgggctcga  
 aagcaaggagtgtgtgaaaaaccttttctcaggttgggaaccgtatgacctggctggg  
 ctacacatgtggatccttccagagtccttgtgtgtggcagcttcttcccagaggctcctcc  
 ggctgggtgagccccctaccaacaacagacaggggggcaaaatatttctacctggacaag  
 gctgccctgagattgtccctttccctcctattaagggaacattacatgcttaagaccttcc  
 50 cagaaaagtacacttcaagggtgacttggctttcatcatgtctgctgacacttaggctcca  
 ctattttaccatgatgggtgtgtgctaacggtccttcccttcccaataacctcaccatcga  
 tggcattttaaatatcactctgttctctgggaccgagggatggagaaccgctctccctca  
 gaccaggttttgaactcaggagctgggttttattttgaagaaacttccctacatgagtcac  
 gagcaaggggaaatggatgtgggggagggaggggctctgagggaggagtacgaatgga  
 55 ggaaagaaaagaatgtcattggcgagggagagcatggcacagcccagggttccctctct  
 tccctccacctccttcttcttcttctcctgcagacggggaactccagtcctctcagatggga  
 actgagttcacccctgggttcccaacgcatacgggttccagcttcgcttctgttttagcatcac  
 ctttctctgtctttatcgtaacattacgcgttgggttcccacggcttctacacactt  
 ccatggccgagaaatggcggttggccatgggcagcaggtccagttcattcttcacagggtg  
 60 ggaagttgtttctcagccaagaagctgatcttctggcacattccaccgtgggtcaacctc  
 tgtttccctttgacctgggtccttttcttccctcctcccttaggaacatcgagttct

>scr\_gb-m59814\_4 (TOXMARKER Assignment: 4; SEQ ID NO: 4)

45 tttaacaatgagacatatacag  
 ctttattttaacctgtaaaaagtcacactctgcagagtgacacctttcttatctcagcaga  
 aagcaaggagtgtgtgaaaaaccttttctcaggttgggaaccgtatgacctggctggg  
 ctacacatgtggatccttccagagtccttgtgtgtggcagcttcttcccagaggctcctcc  
 50 ggctgggtgagccccctaccaacaacagacaggggggcaaaatatttctacctggacaag  
 gctgccctgagattgtccctttccctcctattaagggaacattacatgcttaagaccttcc  
 cagaaaagtacacttcaagggtgacttggctttcatcatgtctgctgacacttaggctcca  
 ctattttaccatgatgggtgtgtgctaacggtccttcccttcccaataacctcaccatcga  
 tggcattttaaatatcactctgttctctgggaccgagggatggagaaccgctctccctca  
 gaccaggttttgaactcaggagctgggttttattttgaagaaacttccctacatgagtcac  
 55 gagcaaggggaaatggatgtgggggagggaggggctctgagggaggagtacgaatgga  
 ggaaagaaaagaatgtcattggcgagggagagcatggcacagcccagggttccctctct  
 tccctccacctccttcttcttcttctcctgcagacggggaactccagtcctctcagatggga  
 actgagttcacccctgggttcccaacgcatacgggttccagcttcgcttctgttttagcatcac  
 ctttctctgtctttatcgtaacattacgcgttgggttcccacggcttctacacactt  
 ccatggccgagaaatggcggttggccatgggcagcaggtccagttcattcttcacagggtg  
 60 ggaagttgtttctcagccaagaagctgatcttctggcacattccaccgtgggtcaacctc  
 tgtttccctttgacctgggtccttttcttccctcctcccttaggaacatcgagttct

catgccattaccgacgggtgactgggttcattctggaccctcatcgagtggatgctgctaaga  
 atcttcttctgatggcctgccaagggtgaccctattctcaggaggtcttctgatgtcatc  
 tgggtgaccagctggagggaggtgaagccagcgggtgaggaagctgtccctgtactggacc  
 attttgatggcacttagccagtcacccaggtggtaaagggcgtgaagtctgggtagag  
 5 cggtaagcaggggttgggaaggcacagcgggtgatgggtgccacagtccttgagactagct  
 ggggtccggatcatcttgtccagggtgttgacgatctctgcaaaacggggccggctattt  
 cgatccttctgccaacagtcacgcatgagctgggtgcagggcagctgggcagtcctatagga  
 gggggcagccggtagtctgtcattggcattgatgacatcttgattggacatatcccag  
 10 taaggtctctctccaaatgacattacttcccacatgacaatcccgtagctccagacatcg  
 ctggctgacgtaaacttgcggtaggcgatggcctctggagctgtccatctaacagggtatc  
 ttccctcccaaggagctgggtgtaggtgggtctgaggtgtcatcctggaggtagcgagag  
 aggccaaagtgcagacactttgcacaccaggttgctgttcaccagaatgttcctagcagcc  
 aggtcccgggtgcacataattcatctcagataggtacttcatgccagcagcgatgcccctc  
 agcatccccacaagctggatcacgggtgaactgtccgtcattttgccggaggaaagagtct  
 15 aaagcgccattctccatgaactccgtaatgatcatgacaggtcggctcttggtgacaaca  
 cctctagggcaatgatgttgggtaggtcaaactggccccatgatgctcgctcgcctcaga  
 aaatcccagcgtgtttctctgagtaccagctttcagggctcttgatggccacatagatt  
 tccctcttgcttggcagcttcaatcgcccttgtagacttctccaaactcccctgctccg  
 atgacctcttcaattttcacaaaagacacatcaatctccttggcaaaactcccggacagct  
 20 tcattagggctctcataagtgaacgggtcaatgtagatcttcatccctggggagcctcgg  
 cctgtgctgtaatgctgaagtttatcactgtacacagcctcttctgctgtaagctcgtttc  
 ctgctgcagacaatggagatagccaccagagacacacaaatacaaccccagctgctgca  
 gagccagcgatcaggggtagctgctctcagctccgacttgtaatcatcatctgtcaga  
 gtctggaagcacatcttgccactgaacttgccatagccagccacgggttcgagctcgtacc  
 25 tggaccacatacaccatgcccggcgtagcccatcgatacgtgcccgtgttggtctggctc  
 ctggccatggaagagttgaactcattgtgctccttctcatagtaccggatctcatagtcc  
 aggatgatgccattaggtgctccggctgagggcatgacaaggtgatgctcctcatggtg  
 gcactgacctgggtgcatgataggaacagtgaggggggcagcttggtttgtggtgatgtg  
 acagagacatgctgtggggggaagggaactcttgctagagactccattgatggcctggata  
 30 tcaaaagtgtatggggtgtggggcccataggctactgatagagacacgacactcagtcaag  
 cccagctgtctgggtacaaactccacattgtcatcgagcgggagcaactccggcgggtct  
 gctctgcaacttcttgagatgatgtttaggtcacatcatctcgcccaccggctctctctt  
 ggaggggtgccactctagaatgatagatgtctcattcacaatggagatgacatttcgaggg  
 cctgatgggacactagtgcacgccacttctgggggatcaaagtctgctcggtaatagcca  
 35 gtccggcaggtgcagatgggagacgcctctgaaggggagcggctgttggaggggagctgg  
 gagcagccttcagcttccctggctggccttgaaggttcccgcaggacaggccttgcaaggcc  
 acgctgttctcaggttcatagcagccttacaggtgcagcgcccaatgggcaccatccac  
 tctccatctccattgcagtagagttttatgggcacatccacttcttctgcattagggatg  
 catgtgccccgagcaatcaccagagatgtgctctctgctcctgtcatggtttctgggaac  
 40 actgcaaaatttgcacaatgctgggacactttttgaagaagacacggacagaaagtaga  
 gacatacaggctccataatcctggaaagcgaggtaaaaaccattcctagtaagaggccca  
 aagctcctgacttctgtgttgaccttcatcaaccttccccaaaaatccacctgggagaag  
 ctctcatctgcagcaatggtgtcaactttgaggtagggggcttcagaccagaaggctgac  
 ttcttgggtggcaatgacagagtcagctctcatagtagtataagttgaaggtctcttgcag  
 45 gagcctgggacatttggaaaggctgctgcagtcctcacagtgaagcgcatctctgtatag  
 atgcgatgggcgccccgtctgttgataaagggtggttaagcagccagttgttctgggttgggt  
 tcaaaagacgttgcacacttggttaagtacggatggtgttcaggttttcatcgtagccactg  
 acttcttcccaccagaggcaggggttgccgtccatcccaactctgcagtggcagtcctt  
 gtgtccatcaatgtttcttccatcgcgggccactgcagatgccaggaggaacagcagcagg  
 50 caatccagggccatcgccggccagcggccccagggccgagccccagcggagacgcgcgc  
 gtcccagggcgccgctgcgctcccggcggggtggcttctccgtgtccttctcgctctggc  
 cgggaccggactccccggagcgcggcgtgggctgggagggagtggtgcgcgcgtggggcg  
 gtgctggcgcgcgctggatgtgggtgtgcatgtgtgtgtgtgtttatgggagaggtggg  
 55 tgtgtgcgtgctgtgtgagaggggtgaggagagcgagccaaaccataaaaaagatgga  
 gggggagtgtgggtgggagaccctgctagtttcatagctggcattcttggggctggaaa  
 ccccatggcacaagacgttaggatggctggtctgctcaaccactgtgccgtgtgtgaggg  
 gtctctcggtgtgtgtctctatcctgctctcattgagtcggatgacctgtacagctctgt  
 ctaccatggaggtatgtattgtgaagtctctgtgctaaggactcacgtttgggtgctttgg  
 60 agatgaaatggatgacatgtacactggatatccccctcgtg

>scr\_gb-m29358\_5 (TOXMARKER Assignment: 5; SEQ ID NO: 5)

5 cccccctcgagggtgttttctttcatttcattccttgtcttttagggctttttttttttt  
 aagggtctcattatattttgttactcttttaagacttatttttgactggactcagattta  
 gaagtagaagctctcagcgaagacagcctacgtctcttggcaatctgttcttggcgcttc  
 tctttggcttcttctcattctcttggccaaaagttagcatattctgcagcctcctccttg  
 10 tttttcttagtgcggttgccttctcagagcaatacgtcggcggttgtgttgaggacacgg  
 ggagtaacaagacgctgaatcttggcgcttgggtcctgggcttcttaccttcttgttt  
 aagggtcttctgacaacatactggcgacatcatcttcttggagagattaaaaagcttt  
 cggattctactagctcttttaggtcccaaccgacgaggcacagtggtatctgtcagtcct  
 ggaatatccttctccttttttacaataaccaagttgagaacactcaggttggcatcc  
 15 acaatgcatcctcggacagacttgcgcttctctcctcagttctcctaggtctataacaa  
 gaatgcccccttactcaaaagcaggcgactctgccatgggtcaaaacgccttgcctcatg  
 ggaaaaacctgtttgtcattcccacgctgatccggaccacataacccttcactcttca  
 ccaagagcatcagcagctacttctgtggccatgcgcttctcatagaacgtacgaagcttg  
 cgttcgtcatccacttctatgagtttctgacagccagtggcaggaaggagatattcagc  
 20 tcatcttgacacagccgaccgcttaggaggctgttaccattctgatgttgagcgcc  
 gc

>aj297736 (TOXMARKER Assignment: 6; SEQ ID NO: 6)  
 20 agttgcttcagtgctcccgggtgctggttagtcacgtttcgtgctgctcattctgccaagat  
 gctgaggaacccagacccaagaccaacaaatggaggaagaggaggtcgaaaccttgc  
 ctttcaggcagaaattgcccagttaatgtccttgatcatcaacacttctactcgaacaa  
 agagatcttctgaggagctcatttccaactcctcagacgctctggataagatcagata  
 25 cgagagcttgaccgacctagtaaaactggactcggggaaggagctgcacattaatctcat  
 tccaacaagcaagaccgaaccctcactattgtggatactggcattggaatgaccaaggc  
 tgacttgatcaataaccttggcactattgccaagttaggcaccaaagccttcatggaggc  
 tttgcaggctggtgcagatatctctatgattggccagtttgggtgttgggttttactctgc  
 gtatttgggttgcagaaaagtgcagtgctcatccaagcataatgatgacgagcagtacgc  
 30 ctgggagcttctcagctggaggatccttctcactgtgaggacagacacaggtgaaccaatggg  
 tcgtggaacaaagggttatcttgcattctaaaagaagaccaaaactgagtatttggaggaaag  
 gagaataaaaagaattgtgaagaaacattctcagtttattggctacccattactctctt  
 tgtggagaaggaacgtgacaaggaagtgcagtgatgatgaggctgaagaaaaggaagagaa  
 agaggaagagaaagaaaaagaagaaaggagctctgatgacaagcctgaaatagaagatgt  
 35 tgggttctgatgaagaagaagaagaagaaggatgggtgacaagaagaaaaagaagat  
 aaaggaaaagtacattgatcaagaagaactcaacaaaacaaagccgatctggaccagaaa  
 tctctgatgacattacgaatgaagaatacggagagttctacaagagcttaaccaacgactg  
 ggaagaacatttggcagtaaaagcatttttctgttgaaggacaattagaattccgggctct  
 40 tctttttgtcccaagacgcgctccttttgatctatttgaaaaacagaaagaaaaagaacaa  
 catcaagttgtatgttcgcagagtttttatcatggataactgtgaggagttaatccccga  
 gtatctgaatttcatcagaggggtggtggattctgaggatctcctctaaatatttcccg  
 tgaaatgctgcaacaaagcaaaattctgaaagtattcaggaagaatttgggtcaagaaatg  
 45 cctagaactatttactgaactggctgaagataaaagagaactacaaaaagttttatgagca  
 gttctcaaaaaatataaagcttggaaattcatgaagactctcaaaatcggaagaagctttc  
 agagctgttgagatactacacatctgcttctggggatgagatgggttctctgaaggacta  
 ctgcaccagaaatgaaggaaaaaccagaagcacatctattttatcacaggtgagaccaagga  
 50 ccagggttgtaactcagccttctgtggaacgtctccgaaagcatggcttagaagtaatacta  
 tatgattgagccattgatgagttattgtgtgcaacagctgaaggaattttagggcaagac  
 cttggtgtcagttaccaagaaggactggaacttccagaagatgaagaggaaaaagaagaa  
 acaggaagagaaaaagacaaaatttgagaacctctgcaaaattatgaaggatattttaga  
 gaaaaaggttgaaaaggtggttgtgtcaaacgattgggtgacatccccatgctgtattgt  
 55 cacaagcacatatggctggacagcaaacatggagagaatcatgaaagctcaagccctcag  
 agacaactcaacaatgggttacatggcagcaaaagaaacacctggagataaaccttgatca  
 ctocattattgaaaccttaaggcaaaaggcagaggctgacaagaatgacaagctctgtgaa  
 agatctggctcatcttgcgtgacgaaacagcactcctgtcttccggcttcagtctggaaga  
 60 tccccagacctatgctaacaggatctacaggatgatcaagcttgggtctaggtattgatga  
 ggatgatcctactgtggatgataccagtgctgctgtaactgaagaaatgccaccctgga  
 aggagatgatgacacatcacgcaggaagaagtagactaggcttcaccagaactatgtgt  
 ttgatgcttaccttcatccttctgataatatattttccatgatttttgtttatttttgt  
 taacctttaaactctgtgtggactgaaaactaggggaaggtaaaaaatttctacatgtg  
 atactgtgatactataggtttgactcaagaggttgatagaacggttgtgtgaagcgttaa  
 tgtaacctacgggtacttgttaactatgggggtctgaaagtgttagctgttgagctggat  
 tccttttagtagaccaaattaagatgacttaagtctcatct

>j00719 (TOXMARKER Assignment: 7; SEQ ID NO: 7)

ttgctcctccttctctcctcgtgggtcttctgttactcttagtcaggggacacccaaag  
tcccgtaggaacttcccaccaggacctcgtcccttcccctcttggggaacctcctgcag  
5 ttggacagagggggcctcctcaattccttcagtcagcttcgagaaaaatatggagatgtg  
ttcacagtacacctgggaccaaggcctgtggtcatgctatgtgggacagacaccataaag  
gaggctctggtgggccaagctgaggatttctctggtcggggaacaatcgctgtgattgag  
ccaatcttcaaggaatatggtgtgatctttgccaatggggaacgctggaaggcccttcgg  
10 cgattctctctggtaccatgagagactttgggatgggaaagaggagtgtggaagaacgg  
attcaggaggaagcccaatgtttggtggaggaactgcggaaatcccaggagccccactg  
gateccaccttctcttccagtgcatcacagccaacatcatctgctccattgtgtttgga  
gagcgctttgactacacagaccgagtcctgctgctgttgagctgttctaccggacc  
ttttccctccttaagttcattctccagccaggtgtttgagttcttctctgggttcttgaaa  
15 tactttctggtgcccacagacaaatctccaaaaacctccaggaaatcctcgattacatt  
ggccatattgtggagaagcacagggccaccttagaccaagcgctccacgagacttcac  
gacacttaccttctgctgcatggagaaggagaagtcgaaccaccacacagagttccatcat  
gagaacctcatgatctccctgctctctctcttcttctgctggcactgagaccagcagcacc  
acactccgctatggtttcctgctgatgctcaagtaaccccatgtcgcagagaaagtccaa  
aaggagattgatcaggtgatcgggtcacaccggctaccaaccttgatgaccgcagtaaa  
20 atgccatacactgatgcagttatccacagagattcagaggttttcagatcttgcctatt  
ggagtaccacacagagtcaccaaaagacaccatgttcggagggtacctgcttcccaagaac  
actgaagtgtaccccatcctgagttcagctctccatgaccacagtaactttgaccacca  
gacagcttcaatcctgaacacttcttggtgccaatggggcactgaaaaagagtgaagct  
ttcatgcccttctccacaggaaagcgctttgtcttggcgaaggcattgcccgaatgaa  
25 ttgttctcttcttccaccaccatcctccagaacttctctgtgtcaagccatttggtccc  
aaggacattgacctcacgcccaggagagtggttggaaaaatacctccaacgtaccag  
atctgcttctcagctcggtgatccggctgaggcagccaggtgccccagttctgttgggaa  
tggcctcatgtttctgctctggtgggacctgctgaaaaccagggtccaaggccactgctc  
cacatct

>j00720 (TOXMARKER Assignment: 8; SEQ ID NO: 8)

cccagtgcccttttgtcctgtgtatctgtttcgtgggtgctcttgccaacatctatggtgt  
gggtaagggaatgaggagtgaatagccaaagcaggaggcgtgaacatctgaagtgcata  
actgagtgtagggcagattcagcataaaagatcctgctggagagcatgcactgaagtct  
35 accgtggttacaccaggacctggagcccagtatcttgcctccttgcctctccttgtgg  
gcttcttgttactcttagtcaggggacacccaaagtcccgtggcaacttcccaccaggac  
ctcgtcccttcccctcttggggaacctcctgcagttggacagaggaggcctcctcaatt  
ccttcagtcagcttcgcgaaaaatatggagatgtgttcacagtaacctgggaccaaggc  
40 ctgtggtcatgctatgtgggacagacaccataaaaggaggctctggtgggccaagctgagg  
atcttctggtcggggaacaatcgctgtgattgagccaatcttcaaggaatatggtgtga  
tctttgccaatggggaacgctggaaggcccttcggcgattctctctggctaccatgagag  
actttgggatgggaaaggaggtgtggaagaacggattcaggaggaagcccaatgtttgg  
tgagggaactgcggaatcccaggagccccactggatcccaccttctcttccagtgc  
45 tccacagccaacatcatctgctccattgtgtttggagagcgctttgactacacagaccgcc  
agttctcgcgctgttgagctgttctaccggaccttttccctcctaagttcatttcca  
gccagggtgttgagttcttctctgggttctgaaatactttcctgggtgccacagacaaa  
tctccaaaaacctccaggaaatcctcgattacattggccatattgtggagaagcacaggg  
ccaccttagaccccagcgctccacgagacttcacgacacttaccttctgctgcatggaga  
50 aggagaagtgcgaaccaccacacagagttccatcatgagaacctcatgatctcctgctct  
ctctcttcttctgctggcactgagaccggcagcaccacactccgctatggtttctgctca  
tgctcaagtacccccatgtcacagtgaagtcctcaaaaggagattgatcaggtgattggct  
ctcacaggccaccatcccttgatgatcgtaacaaaatgccatacactgatgcagtcaccc  
acgagattcagaggtttgcagatcttgcctcaattggtttaccacacagagtcaccaaaag  
55 acaccatgttccgagggtaacctgctccccagaacactgaggtgtatccatcctgagtt  
cagctctccatgaccacagtaactttgaccatccagacaccttcaatcctgagcactcc  
tggatgccgatgggacactgaaaaagagtgaagcttttatgcccttctccacaggaaagc  
gcatttgtcttggcgaaggcattgcccgaatgaattgttctcttcttccaccaccatcc  
tccagaacttctctgtgtcaagccatttggtcccaaggacattgacctcacgcccattgg  
60 agagtggttgcataaacctccaacgtaccagatctgcttctcagctcggtgatcgg  
gctgag

>j00728 (TOXMARKER Assignment: 9; SEQ ID NO: 9)

atggagcccagtatcttgctcctccttgctcctccttggtgggcttcttggttactcttagtc  
aggggacacccaaagtcccgtggcaacttcccaccaggacctcgtccccttcccctcttg  
5 gggaaacctcctgcagttggacagaggggcctcctcaattccttcatgcagcttcgcgaa  
aaatatggagatgtgttcacagtaacacctgggaccaaggcctgtggtcatgctatgtggg  
acagacaccataaaaggaggctctggtgggccaagctgaggatttctctggtcggggaaca  
atcgctgtgattgagccaatcttcaaggaatatggtgtgatctttgccaatggggaacgc  
tggaaaggcccttcggcgattctctctggtaccatgagagactttgggatgggaaagagg  
10 agtgtggaagaacggattcaggaggaagcccaatgtttggtggaggaactgcggaaatcc  
caggagagcccactggatcccaccttctcttccagtgcacacagccaacatcatctgc  
tccattgtgtttggagagcgctttgactacacagaccgccagtttctgcgcctgttgag  
ctgttctaccggaccttttcccctcctaagttcattctccagccaggtgtttgagttcttc  
tctgggtccctgaaatacttctggtgcccacagacaaatctccaaaaacctccaggaa  
15 atcctcgattacattggccatatgtgtggagaagcacagggccaccttagacccagcgct  
ccacgagacttcatcgacacttaccttctgcgcagtgagaaggagaagtccaaccaccac  
acagagttccatcatgagaacctcatgatctccctgctctctctcttcttcttgctggcact  
gagaccggcagcaccacactccgctatgggttctctgctcatgctcaagtaccccatgtc  
acagtgaagtcctaaaaggagattgatcaggtgattggctctcacaggccaccatccctt  
gatgatcgtacaaaatgccatacactgatgcagtcacacagagattcagaggtttgca  
20 gatcttgcccccaattgggttaccacacagagtcaccaaagacaccatgttccgagggtag  
ctgctcccccaagaacactgaagtgatcccatcctgagttcagctctccatgaccacag  
tactttgaccatccagacaccttcaatcctgagcacttcttggtgacgatgggacactg  
aaaaagagtgaagcttttatgcccttctccacaggaaagcgatttgtcttggcgaaggc  
attgcccgaatgaattgttctcttcttccaccaccatcctccagaacttctctgtgtca  
25 agccatttggtcccaaggacattgacctcacgcccattggagagtggttgcaaaaaata  
cctccaacgtaccagatctgcttctcagctcggtag

>100320 (TOXMARKER Assignment: 10; SEQ ID NO: 10)

atggagcccagtatcttgctcctccttgctcctccttggtgggcttcttggttactcttagtc  
30 aggggacacccaaagtcccgtggcaacttcccaccaggacctcgtccccttcccctcttg  
gggaaacctcctgcagttggacagaggggcctcctcaattccttcatgcagcttcgagaa  
aaatatggagatgtgttcacagtaacacctgggaccaaggcctgtggtcatgctatgtggg  
acagacaccataaaaggaggctctggtgggccaacctgaggatttctctggtcggggaaca  
atcgctgtgattgagccaatcttcaaggaatatggtgtgatctttgccaatggggaacgc  
35 tggaaaggcccttcggcgattctctctggtaccatgagagactttgggatgggaaagagg  
agtgtggaagaacggattcaggaggaagcccaatgtttggtggaggaactgcggaaatcc  
cagggagccccactggatcccaccttctcttccagtgcacacagccaacatcatctgc  
tccattgtgtttggagagcgctttgactacacagaccgccagtttctgcgcctgttgag  
ctgttctaccggaggttttcccctcctaagttcattctccagccaggtgtttgagttcttc  
40 tctgggttcttgaaatacttctggtgcccacagacaaatctccaaaaacctccaggaa  
atcctcgattacattggccatatgtgtggagaagcacagggccaccttagaccaagcgct  
ccacgagacttcatcgacacttaccttctgcgcagtgagaaggagaagtccaaccaccac  
acagagttccatcatgagaacctcatgatctccctgctctctctcttcttcttgctggcact  
gagaccagcagcaccacactccgctatgggttctctgctgatgctcaagtaccccatgtc  
45 gcagagaaaagtcctaaaagggttgatcaggtgatcggttcacaccggctaccaaccctt  
gatgaccgcagtaaaatgccatacactgatgcagttatccatgagattcataggttttca  
gatcttgctccctattggagtaccacacagagtcaccaaagacaccatgttccgagggtag  
ctgcttcccccaagaacactgaagtgtaccccatccggagttcagctctccatgaccacag  
tactttgaccacccagacagcttcaatcctgaacacttccctggacgttaacggggcactg  
50 aaaaagagtgaagctttcatgcccttctccacaggaaagcacatttgtcttggcgaaggc  
attgcccgaatgaattgttctcttcttccaccaccatcctccagaacttctctgtgtca  
agcctttgggtcccaaggacattgacctcacgcccaggagagtggttgcaaaaaata  
cctccaacgtaccagatctgcttctcagctcggtag

>m11251 (TOXMARKER Assignment: 11; SEQ ID NO: 11)

cccagtgcccttttgcctgtgtatctgtttcgtggtgtccttgccaacatgtatgggtgt  
55 gggtaaggggaatgaggagtgaatagctaaagcaggaggcgtaacatctgaagttgcata  
actgagtgaggggggcgattcagcataaaaagatcctgctggagagcatgcactgaagttc  
accgtgggttacaccaggaccatggagcccagtatcttgctcctccttgctctccttggtg  
60 gcttcttggttactcttagtcaggggacacccaaagtcccgtggcaacttcccaccaggac

5 ctcgccccctccccctcttggggaacctcctgcagttggacagagggggcctcctcaatt  
 ccttcatgcagcttcgagaaaaatattggagatgtgttcacagtacacctgggaccaaggc  
 ctgtggtcatgctatgtgggacagacaccataaaggaggctctggtgggccaacctgagg  
 atttctctggtcggggaacaatcgctgtgattgagccaattctcaaggaaatattggtgtga  
 tctttgccaatggggaacgctggaaggcccttcgagcattctctctggtaccatgagag  
 actttgggatgggaaagaggagtggtggaagaacggattcaggaggaagcccaatgtttgg  
 tggaggaactgcggaatcccaggaggacccactggatcccaccttctcttccagtgc  
 tcacagccaacatcatctgctccattgtgtttggagagcgtttgactacacagaccg  
 10 agttcctgcgccctgttggagctgttctaccggaggttttccctcctaagttcattctcca  
 gccagggtgtttgagttcttctctgggttcctgaaatactttcctggtgccacagacaaa  
 tctccaaaaacctccaggaaatcctcgattacattggccatattgtggagaagcacagg  
 ccaccttagaccaagcgtccacgagacttcacgacacttaccttctgcatggaga  
 aggagaagtcgaaccacacagagttccatcatgagaacctcatgatctccctgctct  
 ctctcttctttgctggcactgagaccagcagcaccacactccgctatggtttcctgctga  
 15 tgctcaagtacccccatgtcgcagagaaagtccaaaaggagggtgatcaggatgatcggtt  
 cacaccggtaccaacccttgatgaccgcagtaaaatgccatacactgatgcagttatcc  
 atgagattcataggttttcagatcttgtccctattggagtagcacacagagtcaccaaag  
 acaccatgttccgagggtacctgcttcccaagaacactgaagtgtaccccatccgaggt  
 cagctctccatgaccacagtagtttgaccaccagacagcttcaatcctgaacacttcc  
 20 tggacgttaacggggcactgaaaaagagtgaaagcttccatgcccttctccacaggaaagc  
 acatttgtcttggcgaaggcattgcccgaatgaattgttctcttcttaccaccatcc  
 tccagaacttctctgtgtcaagccatttggctcccaaggacattgacctcacgcccagg  
 agagtggcattggaaaaatacctccaacgtaccagatctgcttctcagctcggtgatccg  
 25 gctgaggcagccatgtgccccagttctgttgggaatggcctcatgttctgcctctgggg  
 gacctgctgaaaaccaggct

>m26125 (TOXMARKER Assignment: 12; SEQ ID NO: 12)  
 gacttgggaggaaccaggccctacacttagccctggtaaacagcagagcatgctgggata  
 30 attcttccagaaaaggaaaagcaggcacttctgttcccagggaacacaggagcact  
 ttggacctccctgctgcagtcaggagtcattgtggtggaacttgcctggcttcccttct  
 gggctttgtcatctactggtttgtctcccggaacaaggaggaaaccttaccactaggaga  
 tggatggtgggggcccagggtcaaagccatcagccaaagaagatgagagcatccggccctt  
 caagggtggaacatcagatgaggagatcaaggacttacaccagaggatagataggttccg  
 35 ggcacccccacctttggagggcagccgcttccactatggcttcaactccaactacatgaa  
 gaaagtgggtgctcactggaggaacgagtttgactggaggaagcagggtggagatcctcaa  
 ccagaccctcacttcaagaccagatcgaagggttgacatccacttcattccatgga  
 gcctccccagctgccctcagggcgacccccaaagcccttgcctgatggtgcatggctggcc  
 tggatccttctatgagttttacaagatcatccactactgactgaccccaagtcccacgg  
 40 tctgagtgacgagcacgtgtttgaagtcattctgtccctcgattcctggctatggctactc  
 agaggcatccagcaagaaagggttaaatccggtggccactgagaggatttctacaagct  
 gatgacacggctgggcttccagaaattctacattcaaggcggggactgggggtccctcat  
 ctgcaccaacatggcccagatggttcccacacgtgaaaggcctgcacttaaatatggc  
 tttcatttcgagaagtttttacaccatgactcctctcctgggccaacgcttcgggagatt  
 45 ccttggctacacagagaaggatattcgagctcttgtacccctataaggagaagggtttcta  
 cagcatcatgaggagagtggtacttacacatccaagccaccaagccagacactgtggg  
 ctgtgctctcaatgactctccggtgggctggtgcctacatcttagagaagtctccac  
 ctggaccaagtacagtagcgtgaactggaggatggaggcctggagaggaagtctccct  
 ggatgatctgctggttaacatcatgatctactggacgacaggaaccattgtctcctcca  
 50 acgctactacaaggagaatttggggccagggcattcatggtccataaacatgaggggatgaa  
 ggtctttgtgccactggcttttcagccttcccttccgagctactgcatgcccagaaaa  
 gtgggtgaaggtaagtaccccaactcatctctatttctacatggaacgtgggggcca  
 ctttggctgcctttgaagagcccaagcttctggcccaggacatccgcaagtctgtgctcct  
 ggctgagctgcagtagtgacactggataccaactgtggcttttagcagcagccctggttcc  
 55 tcccaagtcacacttatggaagatgaccccttctgaggaataagtgttccctgacca  
 cactcgaggaccagacttaaaactccacagagtcgtatgttaccctcatatgcttcacct  
 cactacatagctgtgttagctacatggctttaatgataaatggatttatttct

>m34452 (TOXMARKER Assignment: 13; SEQ ID NO: 13)  
 60 tgagccaatcttcaaggaaatattggtgtgttctttgccaatggggaacgctggaaggccct  
 tcggcgattctctctggtaccatgagagactttgggatgggaaagaggagtggtggaaga  
 acggattcaggaggaagcccaatgtttggtggaggaactgcggaaatcccaggaggcccc

5 actggateccaccttctcttccagtgcatcacagccaacatcatctgctccattgtgtt  
 tggagagcgctttgactacacagaccgccaagtctctgcgctgttggagctgttctaccg  
 gaccttttccctcctaagttcattctccagccaggtgtttgagttcttctctgggttctt  
 gaaatacttttctgggtgcccacagacaaatctccaaaacctccaggaaatcctcgatta  
 cattggccatatttgggagaagcacagggccaccttagaccccagcgctccacgagactt  
 catcgacacttaccttctgcgcatggagaaagtgagtcctgcatggatgagagaggagaa  
 gtogaaccaccacacagagttccatcatgagaacctcatgatctccctgctctctctt  
 ctttgctggcactgagaccggcagcaccacactccgctatggtttctgctcatgctcaa  
 gtacccccatgtcacagagaaagtccaaaaggagattgatcaggtgatgtgctctcacag  
 gccaccatcccttgatgatcgtacaaaaatgccatacactgatgcagtcatccacgagat  
 10 tcagagatttgcagatcttgccccaattgggttaccacacagagtcaccaaagacaccat  
 gttccgagggttacctgctccccaagaacactgaggtgtatcccatcctgaggtcagctct  
 ccatgaccacagtactttgaccatccagacaccttcaatcctgagcacttcttgatgc  
 cgatgggacactgaaaaagagtgaagcttttatgcccttctccacaggaaagcgctttg  
 15 tcttggcgaaggcattgcccgaatgaattgttctcttcttccaccaccatcctccagaa  
 cttctctgtgtcaagccatttgggtcccaaggacattgacctcacgcccaaggagagtgg  
 cattgcaaaaatacctccaacataccagatctgcttctcagctcggtgatcggtgagg  
 cageccaggtgccccagttctgttgggaatggcctcatgttctgctctgggggacctgc  
 tgaaaaccaggctcaaggccactgctcacatcttctattgcagttctccaaagtcccaa  
 20 ggcttggttcttattcctgtggaatggcactgaagaagtcaatcgactgtcttattttgaca  
 tgtgaacagagatttcatgagtacacatctcatgctgagtcacttccctcttctctctaa  
 tagccacgtccccacttatcagccctccatggctgtgtgatctgtgtaatggactctgt  
 atatgggtctcagtgctatgtctacagacttacatagtatgtatgggttcaggtaaacagaa  
 tcacagagtgtgtg

25 >u33546 (TOXMARKER Assignment: 14; SEQ ID NO: 14)  
 atggaacctagtgtcctacttctccttgctgtcctcctcagcttcttctgctactcctggtc  
 aggggccaatgcaaagatccatggctcgtcttccaccaggaccctgcctgtaccccttttg  
 30 ggaaatctcttgagatggacagaagaggcctcctcaagtcttttattcagcttcaagaa  
 aaatatggagatgtgttcacagtgcaacttaggactgagggcagtggtcgtgttatgtggg  
 acacagaccataagagaggctctggtggaccatgctgaggctttctctggccgggggaca  
 attgctgggcttgagccagttttccaggactatggtatattcttttccagtggagaacag  
 tggaagacccttcgacgattctctatggccaccatgagagactttgggatgagaaagaag  
 agtgtggaggagagaataaaggaagaatcccaatgttgggtggaggaactgaagaaatac  
 35 cagggagccccctggatcccaccttcttcttccagtgcacacatccaacatcatctgc  
 tccattgtcttttgagagtgcttttgactacacagatcaccaattcctgcacctgctggat  
 ctgatgtatcagacgttttcaactcttaagctcaatcttcagtcaggatatttgaactcttc  
 cctggtgtcctgaagtactttcctggtgcccacagacaaatctccagaaacctccatgaa  
 atcctggacttcattggccagagtggtggagaagcacagggccactttggacccaaatgct  
 40 ccacgagactttatatatacttaccttctgcacatggagaaaaagtcaaaccattataca  
 gagttccatcactggaacctactgtcgtctgtactctctctcttcttctgctggcactgag  
 actagcagcaccacactccgctatggcttctgatcatgctcaagtaccctcatatcaca  
 gagaaagtccaaaagagattgattgtgtgattgggtcacaccggctacctaccctggat  
 gaccgcagcaaaatgccatacacaggcagttatccatgagattcagagattttcagat  
 45 cttgcccctattggaacaccacacagagtcacaaagacaccattttccgagggtacctg  
 ctccctaagaacactgaggtgttccccatcctgagttcagttctccatgatccacagtac  
 tttgaacaaccagacatcttcaatcttcagcactttctggatgccaatggggcactgaag  
 ataattgaagcttttctgcccttctccacaggaaagcgaatttgtcttggtgaaagcatt  
 gcccgaatgaattgttcttcttctcactaccatcctccagaacttctccgtgtccagc  
 50 cctgtggctcctaaagacattgatctcactcccaaagagagtggatttgaagaataccc  
 caagtgtaccagatctgcttcttggccccactga

>x74673 (TOXMARKER Assignment: 15; SEQ ID NO: 15)  
 gaattccgcgccgcccaacgtcctctcttaccgccaaccttcttctgccacctctaccac  
 55 ggtcaccatgtcgcaagccggcctgccactgtgctgggtgccatggagatgggtcgccg  
 catggatgtgacctccagctccgcgtcggtgcgcccttctgagcgcgccacacgga  
 gatagacaccgccttcgtgtatgcgaacggtcagtcctgagaccatcctaggagacctggg  
 gctcggactggggcgagcggtgcaaagtaaaaattgccaccaaggctgccccaatgtt  
 tggggaagactgaagccagccgattctcggttccagctggagacgtcactgaagaggt  
 60 gcagtgccccgggtggacctctcttatttacactttccagaccacggcactcctataga  
 ggagaccctgcaggcctgccaccacgtgcatcaggagggaagtttgggagcttggctt

gtccaactatgtctcctgggaagtggctgagatttgtaccctctgcaagaaaaatggctg  
gatcatgccaaactgtgtaccagggcatgtacaacgccatcaccaggcaggtggagactga  
gctcttcccctgcctcagacacttcggactaagggttctacgccttcaaccctttggctgg  
5 gggcctgctgactggcagatataaataccaggataaggatgggaagaatcctgagagccg  
cttctttgggaatccattttctcaactgtacatggaccgctactggaaggaggaacact  
caatggcatgccttgggtggagaaggctctgaagactacctatggccccactgcccacag  
tatgatctcagctgcccaggtggatgtaccactcactcagagctcaagggcacccaagg  
ggatgcagtcattctgggcattgtccagctctggaacaactggagcagaacttggccttgg  
cgaggaagggcctctggagccagctgttgtggatgcctttgaccaagcctggaacctagt  
10 tggccacgagtggtcccaactatttccgctaagatacatctgccttggggatggcgagct  
tactgcctgccccgccttgctcctgggctcgatctgatctgggttctttccttttagacag  
gtcactgtcctttttcttccctgctttctatacagccagttgctttcaaagtggagagctgg  
ctgagccccaataacctcctgctgaataaaactgttcctgtcacagcctgggctacaact  
ggcggccga

[illegible]

```

>scr_gb-x14254_5 (TOXMARKER Assignment: 17; SEQ ID NO: 17)
40 ttttttttttttttgttctaccttctaccttttattgtcacgtgaacctatggctcctaca
ggctgctgacaagcttggtgagcagggatcccagggcgctcggcaggacatgaggaagg
gttgctgggagggcttggcctcttcttgagaagacagcaaatgtatccagcctagatta
agggtagggcatccctatccctgtcagtgggcctagatctcagagcccacattaaaga
ctgctaattgggtcagaaatgggggtcccttagatgggggtaggcagcaaggccctccctc
45 cagtgttctcattctgttccggttctatttgttgtgtccagggacgggtgaagcagatacc
agtctcaagccccaggggtgcaggaagacgggaaattggggtgtgatgttagggagtgaag
aagggctgaggagcagggagctgccgcgctgcagagctggcttctgtcttcacaagaac
atttgggcccatatcctgcttgggtcactcccaggccagaagatgggtcttccatgtccagt
ggctcactgcagttatggcgccccgcggctcttggtgtgagggacctcagtgccgttgggg
50 aacacacaccagcagtagccagtgctcccatggcactggagtggcatatagttaccggttc
tcatcacacttgggacggaacgccccgggtggacatcagggatgtggctgacttcttcc
tggcacttgggtcaataacttttaggtggagtgctgggtgggctgcttctctccagggagttc
ttgctcatttcaaacacacagccactgtttcatccagctctcaaagaccttccagtcagga
ccattcatagagttctttaagggtgcttcagattctccgggaagctccccctcagctgtggg
55 tagttcacgggtccagacttctgaagcaggtgcatacagtggtcctgggtcatgttgcca
tacttggtaacattcttcacgggcgcttgagcatgttatccatggacagtgggcgcac
agcaagggagtagccatgcgcacgcggctcacagggttggcagatttcggaagcttcatg
cgaaggttctccagttgcaggttctgggaggtgacggtcagcttgtccaggcggccctgc
tgctggtacaggaagtaagcagtggtggcctgccagccaagagcagagccaccaggaca
60 gagacactgggtgacaggactccacgggttgcaattgcttcttggggctctagcacgctgg
cccaggatgggcagctgctcatggttagagatgaggtgcgcgctggctatccatgacteta
gcctctagcttttcccccaagtgcgtgctggtgctgctgctgctgctgctgctgctg

```

84



>scr\_gb-bi275638\_1 (TOXMARKER Assignment: 18; SEQ ID NO: 18)

5 cggcacgaggcgcgctcggcgctgtcagttcgtcccgtgcccctcgcccttgctgctg  
gctctgacggcgacgcagggcgggcgggcccggttcgcgccgagcgggcgccggtgag  
ggcgcgaggaggcgacagcgaggaggagccgtgagcctggcacggagcgcccgcg  
ccatggcgtagcctatctctcaagtacatcatcggcgacacaggtgttggtaaat  
cgtgcttattgctacagtttacagacaagaggtttcagccggtgcatgacctcacaattg  
gtgtagagtttggtgctcgaatgataaccattgatgggaaacagataaaaactccagatct  
10 gggatacagcagggcaggagtcctttcgttctatcacaagggtcatattacagaggtgcag  
cgggggctttactagtgtatgatattacaaggagagacagttcaaccacttgacaacct  
ggtagaagacgcccgtcagcattccaattccaacatgggtcatcatgcttattggaaata  
aaagtgacttagaatctaggagagaagtgaaaaaggaagaaggtgaagctttgcacgag  
agcatggacttatcttcagaaacttctgccaagactgcttctaattgtagaggaggcat  
15 ttattaaacagcaaaaagaaatttatgaaaaaatccaagaaggggtctttgacattaata  
atgaggcaaacggcatcaaaattggccctcagcatgctgctaccaatgcatctcacggag  
gcaaccaaggaggggcagcagggcagggggagggtgctgctgagtcgctgttgccggctag  
ctgccagtgaggccacgcactctgtcaccctctctcctcatgctcagctgagacatgaa  
actattgaaatggctttgtgtcacaggagactttaatccttcagattcttgtataacttt  
20 gaataaatgggttaattgttcaacttaaaaagacagattttggagattgtattcatatctatt  
tgcatttgatttctaggtcaattg

>scr\_gb-x66871\_3 (TOXMARKER Assignment: 19; SEQ ID NO: 19)

25 ttcta  
aagtaaaaatggtttattcacgacacatatgaggaagtgtctcatgtcacagacggtacg  
tccaactccctggaatgttcatttctttggcataaaggagagaatgaggggaaagccagg  
caaaggcagctaagatggggatgggtcggcagctctgtcgctcatcttcacagggaggag  
ttcaggggtccattagtggcaggctgatttctctagaacattaggttggggcacaggtagg  
gccacttctgggcaatccaccatgccaaagccctcagtcgtcccaccacacagggtacag  
30 cagcgcttctggttagtcacccttagtgcttctgctggatgaagtagtacagggatttgcc  
atatttctcttggaattcagatctgattttcaacatgtccacttcactgcgagagaccat  
gattctaatacaggaccttgtctcgagtccttgccttcatggagtacacagccggtc  
agcaaagtacaggggcttgttctgaatgcactgaaccagggttcaggaaggcgttctccag  
gtctcctttgacctcttctctgatgctctccagcatgtcataaggactgtagctcttgta  
cctttcgaacacttcttgagggtggcacacactgcgctcagtcagtgctgatccactt  
35 ggggacatcggttcccttctcttcccccagcatcatagagctcccgggcatcctgggtc  
aatcagctcgtagtcaataacagaaccatcctctgcccgtttaccctttgcaagggcgag  
caacagcttctcggaattctccagatgtgtcagagatgatgtccttctccagatcggtcct  
gtacatttcttatacactcggttaattctcctgcagctcctgggttggttcttgagcagat  
gatctcgatgagggagtctcatcagtcctccaggcccttcatggaggctttgagctcaga  
40 ggcactgtagtgagcaggtgtcttcaacaggcctaacatcacgggtctccagggtgaccaga  
caaggccgacttcatcgccgatggcagttcctttttgggtcctcctctggtaggcgaaggc  
aatgtcctgcctctgtgcattgctgcggttagtcagaatggtgacaatgggtgacctcgtc  
cacgcctttgggtcttgattgctgtttcaatgttcaaagcatcctctcagcgtcgaagtt  
ggtgtagggtttgaccgaccataggcacttgggggtgtagaatgctgagaatcacccctc  
45 caagctgagcttgacagggatttcgtggacagtagacattttgaaaaaaaagctgggccc  
ggcacctattgcagagagcctcc

>scr\_gb-149379\_3 (TOXMARKER Assignment: 20; SEQ ID NO: 20)

50 gggatgacatagagtacaacattcagagaagtttaactattaagtgcgcaggatgaaaggt  
caggaggcaggcctttaactgggctgtgagaatggagaaagcacggtgcactttaacatc  
tgctttcccagaggaaaaagtaaggagaaacagtacaatcatagaagagtcttcgtaac  
agaagcgcgaggagagcattatggacaagttctgcaactctactttttgggattctctcat  
tactggaaagtcagagggtgacctgcctcttggttttgagcaaaactgttctgggtgga  
55 ttcccttgggcttcttgggtcctgggtccttggcaactttacagcgtgtacagatcca  
ggaccaagagatcttctataaccaaattctaccttgccaagcaggtgttcgtcgtgtttc  
ttcttatttttagcagccatagacctgtctcttgcgctcacagaagatactggacaagcca  
cagttcctcctgtcagatatacgaatccaatcctctacctgtgcacatgggtcctgggtt  
tggcagtcacgacagcaggcaatgggtgtgtacgaaagaactcttgggtcctgtctctgt  
tctggatcctctcggtcttatgcggcgatttccagtttcagactctgatcagagcactcc  
60 tgaaggacagcaagtcacaacatggcctactcctacctgttcttcgtctcctacgggttcc

agattgtcctcctgattcttacagccttttcaggaccaagtgactcaacacaaactccat  
 cagtcacggcttcttctgagtagcattacatttagttggtatgacaggactgttctga  
 aaggttacaagcatccactgacactagaagatgtctgggatatcgatgaaggggttataaa  
 caaggtcagtcaccagcaagtttgaggcgccatgacaaaggacctgcagaaagccaggc  
 5 aggccttttcagaggcgctgcagaagtcacagcggaacctgaggccacactacacggac  
 tgaacaagaagcagagtcagagccaagacgttctcgtcctggaagaagcgaaaaagaagt  
 ctgagaagaccacaaagactatcccaaactcgtggttgatcaagtcctcttcaaaact  
 tccacgtagtgatcctgaaatcatttatactgaaattaatacatgaccttttggtgttct  
 tgaatcctcagctgctgaagttgctgatcggtttctggaagagctctaactcatacgtgt  
 10 ggtttggctatctgtgcaatcctaagtgttctgctgactctcatccaatcttctgccc  
 ttcagtcctactttcaacattgttttgtgttgggaatgtgctgacggacaaccgcatgt  
 cttcgatatataagaaggcattgacctatctaaacttggttaggaagcagtaacaccttg  
 gagagacggtgaacttgatgtctgtagattccagaagctaattggatgcgaccaactaca  
 tgcagttggtgtggtcaagtgttatacagattactttgtccatcttcttctgtggagag  
 15 agttgggacgctccatcttagcaggtgttgggttatggttctcctaataccagttaatg  
 gagttctggctaccaagatcagaaatattcaggtccaaaatatgaagaataaagacaaac  
 gtttaaaaatcatgaatgagattctcagtggaatcaagatcctgaaatactttgctggg  
 agccttcatttcaagacgaagtccagggcattcggagaagaagaactcaagaacttgctgc  
 ggttcggccagctgcagagtcctgctgatcttcattttacagataactccaatcctggtgt  
 20 ctgtggtcacattttctgtctatgtcctggtggtgatagcgccaatgttttgaatgcggaga  
 aggcattttacctccatcacctcttcaatatcctacgcttccctctgtccatgcttccca  
 tggtagcctcatcgatcctccaggccagtggttctgtggaccggctggagaggtatttgg  
 gagagacgatttagacacatctgccattcgccgctgagcaattttgataaagctgtga  
 agttttcagaggcctcttttacttgggacccgacttggaagccacaatccaagatgtga  
 25 acctggacataaagccaggccaactggtggctgtggtgggcactgtaggctctgggaaat  
 cctctttggtatcagccatgctgggagaaatggaaaacggttcacgggcacatcaccatcc  
 agggatccacagcctatgtccctcagcagtcctggattcagaatggaaccatcaaagaca  
 acatcctgttttgggtccgaatacaatgaaaagaagtaccagcaagttctcaaagcatgcg  
 ctctcctcccagacttggaatatattgcttgaggagacatggctgagatcggagagaagg  
 30 ggataaatctcagtggtggtcagaagcagcagtcagcctggccagagctgcctatcaag  
 atgctgacatctatattctggacgatcccctgtcggtgtggatgctcatgtgggaaaac  
 acattttcaacaaggttgtgggccccacggcctgttggctggcaagacgagaatctttg  
 ttaactcatggtattcacttccctccccaaagtggatgagattgtagtcttggggaaaggca  
 ccatcttagagaaaggatcctatcgtgacctgttgacaagaaggaggtgttctgtagga  
 35 actggaagaccttcatgaagcattcagggcctgaaggagaggccacagtcataatgaca  
 gtgaggcggaagacgacgatgggtgtattccaccatggaggaaatccctgaggatg  
 cagcttccctggccatgagaagagaaaatagtcttcgccgtacactgagccgcagctcta  
 ggtccagcagccgacgtgggaagtcctcctcaaaaactccttgaagattaaaaatgtgaatg  
 tcttgaaggagaaggaaaaagaagtgggaaggacaaaaactaattaagaaaagaatttgtgg  
 40 aaaccgggaaggtcaagttctccatctacctgaagtatctacaggcagtaggggtggtggt  
 ccatacttttcatcatccttttctacggattgaataatgttgcttttatcggtcttaacc  
 tctggtgagtgcttggaccagtgactctgacaacttgaatgggaccaacaattcgtctt  
 tcatagggacatgagaattggggcttcttggagctctgggattagcacaaggatatgtt  
 tgcctatttcaactctgtggagcatatatgcttgcagaaatgcatcaaaagctttgcacg  
 45 ggcagctgttaaccaacatcctccgggcacccatgaggttttttgacacaactcccacag  
 gccggttgatgaacagattttctggtgatatttctactgtggacgacttgctccccaga  
 cacttcgaagctggatgatgtgtttcttggcatcgtggcactcttgtcatgatctgca  
 tggccaccccagctcttcgctatcatcatcattcctctcagcattctttatatttcggtgc  
 aggttttttatgtggtacttcccgcagctgagacggttggtattctgtcaccaaactctc  
 50 ccatctattctcacttcagtgagactgtcacaggtttgcccattatccgtgcctttgagc  
 accagcagcgatttctagcttgggaatgagaagcagattgacatcaaccagaaatgtgtct  
 tttcctggattacctccaacaggtggcttgcaattcggtggagctggttggaacttggt  
 tctgtctctgttccgccttgctgctggttatttatagaaaaacttaaccggggacgttg  
 tgggctttgttctgtccaacgccctcaatatcacacaaacttgaactggctagtggagga  
 55 tgacgtcagaagcagagaccaacattgtggcagttgagcgaataagtgaatacataaatg  
 tagagaatgaggcgccctgggtgactgacaagaggcctccggcagactggcccagacatg  
 gtgagatccagtttaacaactatcaagtgcggtatcgccggagctggatctggtactga  
 aaggtgacttctgtaacatcaagagcggagagaaggtcgccgtagtgggcaggactgggg  
 ctgggaaatcatcctcacaaactgcctcttcagaatcttagagtctgcggggggccaga  
 60 tcatcattgatgggtagatgttgctccattggactgcacgaccttcgagagaggctga  
 ccatcattccccaggaccccattttgttctcggggagctctgaggatgaatctcgacctt

tcaacaaatattcagatgaggaggtttggagggccctggagttggctcacctcagatcct  
 ttgtgtctggcctacagcttgggttgttatccgaagtgcagaggggtggtgacaacctga  
 gcatagggcagagggcagctcctatgcctgggcagggctgtgcttcgaaaatccaaaatcc  
 5 tggctctggatgaagccacggctgcagtggtctcgagacggatagcctcattcagacga  
 ccatccgaaaggagtctcccagtgacgggtcatcaccatcgctcacaggctgcacacca  
 tcatggacagtgacaagataatggctcctagacaacgggaagattgtcgagtatggcagtc  
 ctgaagaactgctgtccaacagaggttccttctatctgatggccaaggaagccggcattg  
 10 aaaatgtgaatcacacagagctctagcagctgggtccgtggctggcggactataagaaca  
 gtttctattatttggcttgggttctgtgactgtgctctaggtgcaaagacacatatatttg  
 ttcccgttgctcaggtggcctcaaactctaaggctccagcaatctctggctcagccag  
 agacctgtaaaaatagacacttcaaagattatcatgaataaatatttaataaataagtaa  
 aaaaaaaaaaaaaaaaaa  
  
 >scr\_sc-132690501\_1 (TOXMARKER Assignment: 21; SEQ ID NO: 21)  
 15 gaattctctgggcccattccgttgttctcaatggacatgacctccaggaagctaaagtcca  
 ggtcgtgaccaaagccaagggttgtagagcgggaatctgccccggatagcgttgccggacat  
 tcttgaggatctgggaacgggtccgtctccccctcagtgggctctccgtcggtcaacatga  
 taagaattgaggcagggtgctgagttctgggtggcttcccttgagctctgtttaagatct  
 cgattcctcggagcaagcctccattcaggtttgtggctccagccaaagaaaagcgcctca  
 20 caaagtcttgggtgcttggcaaatggcgtgagacgggtaccagtgagcccttccatg  
 actgcacttgagacccaaagaggaccaggtcaaagttgtctactggcttcatgtcccca  
 atatcttaaggagcgcctccttgtctgcttcaacttctggccttccatggaccactga  
 tatcaatcacaaaaaccaggttcttgtctatgttggtcaggttttggggggcaaagaaat  
 gtgtaaagtaattgttggccaccaggaggtcacagagcttgtctcggttcacatcgtagg  
 25 tcaccttgaagtctccattcagcaaggaggttagagcacgtggggcaggactgctgctggc  
 tcacagtggggcggaagagcacatgaccttcttccccgagaaagacttcttgatgggtt  
 ggcacttgactggtgatgacgtagtgggcaaagcgagaggtgactttgcaattg  
  
 >scr\_gb-aw141735\_3 (TOXMARKER Assignment: 22; SEQ ID NO: 22)  
 30 ttttttttttttttttttactgtatatgtaatttaattcaaattggaacaatgacgtaga  
 tatataagccacaatccatgaaagtcttggaggaaaacataggagcagttatttctgtac  
 ttgatttttagtggtgagattcttagctgtggcatggatacacatgatcagaacagtatta  
 aataaggagaacgtcactgaaaagagcaatctgtgtgcatcaaagaacattatcaagaaa  
 gcaaagaagcaatgtgtataaaaacgtccctaataaggtaaatctacatagataaagagaag  
 35 attgggtgggttagacaaccagagggagggaagaatggagagtcactgagtaatgggtacagt  
 gtgtttgaaaggggataaagataagatcgtggcctgattttaccataaattgttgattc  
 tttacacaagaataatgggttagaggaatgagccacaatagcagatattatccaaccatta  
 atgaaacttatgaccacttcttaaatttttattttttttttttttttttttttttttttt  
 40 cataactttgagtgtgtacatgcttatacaggatgctggggccagtagtagccaaata  
 aaggcatcaagacatgggtggaaactggaatttccagaggttgtaagcagccatgtgggt  
 ggtgggaaatgtccctgtgtcctttgcaagatcagcaacttttccctagtatctgtccttc  
 tctccagcattcttacacattgattcagttctaccaggtgtaagttattggctataagt  
 tatgagtatcagcggcatagcaaaggctatatggcatcattagacataacctgcaaagg  
 45 gcacaaatgcattcaggatagggagagctgaatgcaggcatcataagatcaggctggcag  
 gaagaaagtatcctcatcttggaaacatgggttccccctacttgccatcctgacagagct  
 ttggagtggtggagatactgaagagaggactctccccatgtagtaaatgtgtctttatgg  
 agatgagaacctgccacagaacagaatgtgctggttttgtgtgcttgatgaagaaaag  
 gaaggggtggtcagcacagaatgttgggacaaaagcagcacagcagtagttctatgacagc  
 50 ggaggtgctgcagcctctgtgccttccctcattgacctccactacgctcttgtgaacaat  
 ctggacacacacaggtttctctctggagacattgctgataagttagccttggcctcttg  
 gaagactccactattcccaagcgtgaaacacagactccatgtcataatcctcttgcag  
 tttaaatttttggaaaggaaaacctcaacattagtgttcttcataaagtctgggttgggtcca  
 ggctgttaacttctcaaaagtgagattgctttccaccttgcagaggtccccgtcattatc  
 55 tgggagtaggaccacgaagctcagctccattccttcatatggcatcatgagcacttgccg  
 ctgcacctcggtcacatgggcaaggttatatgtgtcctcacacacatcatctgcactag  
 t  
  
 >af184983 (TOXMARKER Assignment: 23; SEQ ID NO: 23)  
 60 gtatttcataaaacagagaggatcgagaggccggcactctgactcctgggtggatggga  
 ctaggagtcagagtcaggccctgactggctgaggcgggcgctccgagtcagcatggaa

5 agtctctgcggggtcctgggtatttctgctgctggctgcaggactgccgctccaggcgggc  
 aagcgggttcctgtagtgctggtggccatgagcagtatccggatcacatgagggagaaaca  
 caattacgtggctggtcttcagatgaaaatgaatgggatgaacagctgtatccagtgtgg  
 10 aggaggggagagggcagatggaaggactcctgggaaggagggcgtgtgcaggcagcccta  
 accagtgtatccacggccttgggtgggttccaatatcaccttcgtagtgaacctggtgttc  
 ccagatgccagaaggaagatgccaacggcaatatcgtctatgagaggaactgcagaagt  
 gatttggagctggcttctgacccgtatgtctacaactggaccacaggggcagacgatgag  
 gactgggaagacaacaccagccaaggccagcacctcagggtccccgcagcgggaagcccttc  
 15 cctcgccccacggacggaagaaatggaacttcgtctacgtcttccacacacttggtcag  
 tattttcaaaagctgggtcagtggttcagcacgagtttctataaacacagtcacttgaca  
 gttggccctcaggtcatggaagtgtattgtctttcgaagacacggccgggcatacatccc  
 atctccaaagtgaagacgtgtatgtgataacagatcagatccctatatctcgtgacctg  
 taccagaagaatgaccggaactcgtctgatgaaaccttccctcagagacctccccattttc  
 20 ttcgatgtcctcattcacgatcccagtcatttccctcaactactctgccatttccctacaag  
 tggaaacttggggacaacactggcctgtttgtctccaacaatcacactttgaatcacacg  
 tatgtgctcaatggaaccttcaactttaacctcaccgtgcaaaactgcagtgcggggacca  
 tgccccctcaccacaccttcgccttcttcttcgacttctccttcgcctgcactcttcgct  
 tcacccacattatcaacacctagtcctcctttaaagcctactggctacaaatccatggag  
 ctgagtgcatttccaatgaaaactgccgaataaacagatatggttacttcagagccacc  
 25 atcacaattgtagatggaatcctagaagtcaacatcatccaggtagcagatgtccaatc  
 cccacactgcagcctgacaactcactgatggacttcattgtgacctgcaagggggccact  
 cccacggaagcctgtacgatcatctctgacccacactgccagatcgccagaacaggggtg  
 tgcagcccggtggctgtggatgagctgtgcctcctgtccgtgaggagagccttcaatggg  
 tccggcacgtactgtgtgaatttcaactctgggagacgatgcaagcctggccctcaccagc  
 30 gccctgatctctatccctggcaagacctaggctcccctctgagaacagtgaatggtgtc  
 ctgatctccattgggtgcctggccatggttgcacatggttaccatcttgctgtacaaa  
 aaacacaagacgtacaagccaataggaaaactgcaccaggaacgtggtcaagggcaaggc  
 ctgagtgttttctcagccatgcaaaagccccgttctcccgaggagaccgggagaaggat  
 ccactgctccaggacaagccatggatgctctaagtcttcaactctcacttctgactgggaa  
 35 cccactcttctgtgcatgtatgtgagctgtgcagaagtacatgactggtagctgtgttt  
 tctacggattattgtaaaatgtatatcatggtttaggagtgtagttaattggcatttta  
 gtgaagggtggaagacagtatttcttcgcactctgtattgtggtttttatactgttaat  
 aggggtgggcacattgtgtctgaagggggagggggaggtcactgctacttaaggtcctagg  
 ttaactgggagaggatgccccaggtccttagatttctacacaagatgtgcctgaaccca  
 gctagtcctgacctaaaggccatgcttcatcaactctatctcagctcattgaacatacct  
 gagcgctgatggaattataatggaaccaagcttgtgtatggtgtgtgtgtgtacataa  
 gatactcattaaaaagacagtctattaaaaaaaaaaaaaa  
  
 40 >scr\_cg-22510674\_1 (TOXMARKER Assignment: 24; SEQ ID NO: 24)  
 gaattcttgcagttacagagtatggctgttgtctactcgggagctcccagatcctcataa  
 ctcagggaagctgtccctatttatggacaaaaaagtttgacgccaggtcgggcctacatga  
 gctcttctctacctgcaagtccccagtgatctgaggaagggtgattctgtcagagaag  
 caaggaagatcaatgcacacctttagtctcagcccataggaggcagagtcaagcagatc  
 45 t  
  
 >scr\_cg-57215224\_1 (TOXMARKER Assignment: 25; SEQ ID NO: 25)  
 aagctttatagtcaggcacagctggctgttgccaggcaactgtggggcagagcatacctg  
 gctgttgccaagtagctgtggggtggagcttagacagaatcccaacagatagtatagttg  
 50 gagaggggtttcagctctgtcacagtgaggagggcaggggcagtagttgagttcatggtgacc  
 agatcttgtgatggaggaaatttacatcatcatcccaggctagaaagcagtgagcagggc  
 agagacaggagcaggttatcaccttgggaagacctgacactagt  
  
 >scr\_gb-aa850767\_2 (TOXMARKER Assignment: 26; SEQ ID NO: 26)  
 ttgcgcccgcccaagtctgccacttcaacactgtatctaaaacttgaaaggcactgtcaa  
 55 aaacctggtgggttccctagctttagggtatccatcgtagagtcagtaaacatggcaact  
 ctgcctccgggcatgtgatacgtcgccagcagaggcttgcctagcccttgccacacaacgc  
 tcagcttactcaaagcactgccaagacatggctgccttgagacgggtgtctgggtcctt  
 ccttccataccttagggcgcccccttcacagcactgggttaagcaatcagccccctcccg  
 agaggagaagggaaggtaaaagacaaaggtatgttttactatgcaaaacgttccagag  
 60 ggggaagatgaacgaagtaacaagtatccaacacagggttttaaaagcaacgacatttc

aaatgagcttgtatgggagaaagaaagcagggttttcaggaaaaatccaaacacattcag  
gtgtgtcttttaagtcaggtttatcatttattctaagttcattgggaggaactgga  
gactatcagcatagctgtcttactgggaaggcattcccagtgataaacatctccctta  
cctgagctcttggcgagagattctgccagcttgactctctc

5

>scr\_gb-ai011994\_2 (TOXMARKER Assignment: 27; SEQ ID NO: 27)

ttttttttttttttttccagaaatttgccattcttttatttgaaggcaaaaattccca  
tggaagtctggatgaagagagagacaaaggcttatagaaaataaattgaataactagaga  
ttctctggatccagacatagttggttgataaatttgttacctatttctcattgtatttca  
cattatttagacatagttcttgacatctctgttttgatactgtctctggccaagagttt  
tggctcttcttctaaatatcaagaggaaaaatggcagaacaaaccagtaattgttacatg  
gcagtgtggttctctgagtataataatcaagcatttagcagcagttgtagttatctgaatataa  
tgcatagatataatacatgaccgaagagacacaccgatttaacaaccaatgtcaacact  
gaaacaaagaattttaatgctaaggcacccaatcacgggtgtctttcagttatttgttggt  
ttcttttaggagactggccatacacagcagggattcaaaattgtggcttgagtcagtcgaat  
caacatttgcatcttgagtaacttaccatctcttcttatgtctccacaaacatagtttcag  
ttgggataatcactgaggtgtgcacagcccttcttctgtagtttaggcaatatccaag  
gctgtagaacttggggtaagggtgtaattggtgtcacaggaggagacatctactcactgtta  
aatgttgctctctgatgtagggtggccatagctcccatacagatctcacagggaagccgatg  
ggtaatagcagcaggaagatctgggtctacataactgactctggaacttcttgacttata  
acttattactttttgggtttcttttc

10

15

20

>scr\_gb-aw142293\_1 (TOXMARKER Assignment: 28; SEQ ID NO: 28)

agggaaacccggtttctgaggttaagaacctggtatgaggtagaaagcagaatcggacctt  
aggcactcgagcgtcgtgtcgaagaacattaaatagaatagaggagtaaaggggatggt  
tcggataagcgtcaggtcgagtcgaagaagtccttgcaagaagagtttaaggagcaagaat  
ttctagaagcatctagataaggagtcgtagcactgacgttactagtaataagtaggggt  
gagtcggagaatcatgcgctcgatggtcataagatagtagtatctatcgaggagtgtagggg  
cctcgctccttcggcggaagaagtaacgcgtagcggtaagaatcttgctgcttcattatctt  
aagggttaaggagccatcagtttagaagtcgttcccgcggtagtaagttcgcgtcgatttt  
aataagacttttagattgcgtcggttagtcgacgtagtagacgggttaatagtaacggtcct  
acttccttaagcgtttcgctagttcttaagcttaattcggctactctagattttaccttt  
ggggtaagtttccggttagcgttggtggaatcggttttgctgcgggggtggacgcccgtc  
taggagaacgcattcgctacgaacggtgc

25

30

35

>scr\_gb-bm383327\_1 (TOXMARKER Assignment: 29; SEQ ID NO: 29)

tttttttttttttttttgatggccagtgacagtttttgcttttttttatatttataaaca  
aaccaacctccccccaagtaactccccaaacaaaaaaaccagattaaataaaattt  
acagtgaacccagcaaacatctgtatgtgcaattaaatactgtgtctgttactgtgggtgg  
cacgaacctcaaacaaacaatatacaagtggtctgggggttgatcaggggtcgggggagt  
ccaagttttaactctgtgggggttgaggagacaaggtggggaattgaacgaatggggaa  
atcaattttatttttcttaattctgtccatataaatatattcatgaagacaaaagaggga  
agggcagttgggctggtgatgaagtgggagaaggggagggcagagccctctcaactctac  
tcagccaaaaatatgaaacaaatttaatttcatggtgggagaagagatttaaaaaatgata  
gaagatgggaaggagggggagacagaaggggaccaaccagggaaaagggggacccatggc  
aaggagtgccatgtcaaggagtcctgtgccggtgtgagaatctgtctgcttctctcttc  
agccataatgtggttaagctctggcccaatccgccttcggctcccggcttgcccttgctc  
ctattgtgccagccctcccgcctccagctattgagagctagctcgctccaggatcctca  
ggctcgtagttcttttagctactcgaagtttgaagcgactcacagagttggtgaggcgaa  
gggaggcattgtgggcagccaggggactggggaacacagccactatagtgtaacaggcag  
cgaggtccgcagtgccggtctctcagcagtcctcactgttggtccccccacctgcaccag  
gcaacccctgagcatccttaagccactggatcttggcaccagacatggcaagctgtgtga  
agagtttgctcctctgtgcgggtgattccttcggggagatcagtcacctccagtaacc  
tcccagcacaacatccgctgtcccaggtcagtgaggcagacttgagtgcttgctctct  
tgectcggtttccatgcttcaatccactctgtccctggtgcaccgtatacgttgactggc  
catgg

40

45

50

55

>scr\_gb-bm386625\_1 (TOXMARKER Assignment: 30; SEQ ID NO: 30)

ttttttttttttttttcacatgtcaacaactgctagctactattaaaatactgtcaccc  
aaggaggtggaatgtttaacagaaaataggctttaacaattcatactggctcctcaataac

60

5 tgcagatgactagttcaagccaactgcaaaactgagcaagaaatgcagcttgaagaacag  
 gacaataaaatttaattcttgcaacttgatagacttggaggcattccggtcaatgtagaag  
 accttgccgggctcagagttaaagcccaggccagccccctaggtgtacttccagctcatg  
 gcccggctcgtagtcctgctgcagactctgctggagcgtatctgaagacttctgtccagg  
 gccatgttgacctgacagtcctgctgggaggacggttgaatgacggggatagatgctta  
 aagccgcccataagtttcaggaaatttttagtttctgtttcttcattttcaaagcccgcag  
 tgtcccactggccaaactgggttccctgatctatgcaggcctcatccatattgccttttt  
 tttccagtaccacctccaagtctgtgtctgactcctcttttttctcctgcccagggtctt  
 10 cctttactccgctctctttttctcctcttcttttcttcttcttccacagccagccctcac  
 caactggctgctccaccttttttttgatttcttcttcttcttccagggccttcaggc  
 tgtctatcgggatgaaatccaacgcttcttctgactgacttcttattccctttcttca  
 agctgttctccatggagatcttggagttgaccgggaggatgtctccctcctgatgagtct  
 ttttcttcttcttctttttcatgctgtgatctctggggctcccctgcttcttttgtgcc  
 ccaaggccgctgctcttcggcctctgccccctctaagcatgagtgcaaagcatccccag  
 15 cctcagggatccaagagtcctggggaggaaaagcttccatgtccggaagcttcttctcct  
 tccctgtgcttttttaggcttccctaccagcttctgcttacctccttggcatgcttggagtcg  
 gggaagtcttcagccctgagccttgggaggcaacctttgataaggacttccgctcttct  
 ttttcttctctctgatgaggccttctgctgactgctcaaggacctgccttctaaggctag  
 gtgactttatctgtcttgtccatgtaggctcatccttggccaggactcatccaagtgtg  
 20 tgctacaggacttttttcttcttcttcttcttggccagtgcacatctcaggacctgcacct  
 caccacattgttaaaaggggatgtagcccttggaggagaaacatctatgaagtaacat  
 tattgtttaagactgagtactgagtctctggttctgagacatttgccaccttcttttct  
 tcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct  
 25 tgaccattactggctctattagcaggccaagcatccccacgtgagcaccgcgcagccgcg  
 acccggaagtgcagcttcgaatttctggccgccccctcgaaatcggttctccttccgggtc  
 gcagcttcgcgcgcgccttgggttctgctgtagaaacggcgctccatggccgtgcctagacaag  
 catccagcctcagcgtgctgcgtgaggagacgggagggcgtgcggactcgcgggtcacta  
 cacgaatgccccgggtcgcagggtcgcctgggtcccccgaaagtctcgtgttaccgcgc  
 aggtcgcgcgagcctccggggaagaacctgtgggagcagatctgcgaggagtatgaagccg  
 30 agcagcctaccttccggaaggatataaagtgaagtttagtttccctgccttgcgcggaat  
 gctacgctttcacgtggccatcttccccgcagttgttgacatgcctagtgcacctgacct  
 ctgacaccgcttttccaccttttggcaggatctgtattttaacttacttcagagtcctct  
 tagttgtcttgggttgggggtgggttgggggtgttgggataacagatggggcaaggctgt  
 agccctactgagctgtttccagaggcgttgtcaggaaggatttccagtgttacagcccc  
 35 agagtataaacagcagcgcctgttagcttaatgggtccccattgggttctgtggctgcggct  
 caccaggatttctccattcaaaaggcccagacatggctgacagcctcctctgtaggctctg  
 actgacaagctaccacgcgtcttaggtaaatagtaaagcctttattttctgttaagaac  
 agcattttgaaaataaaacctatctgcccagcttaacaaccttttaaagtctgtgatatt  
 ttatatacagccctgtacatactgattgtctggaaatttcttaaacagtttttgtttata  
 40 agtatgcaagtcagccaggatgaggggaagagtgaggggtacattataaaaatacacattaa  
 tacatttaataaaatataattatctatcaaaaacgagccatagctcttaataaagc  
 acctgccagggtctctcatcagctcacagttgtacatccttggatgtgtaaatgccagt  
 gcccccttctactttgccatttggcgaattcaaaagacaactcttccaccacctgcact  
 45 tgttccctggccttgacctcctctgtgtgggggtggggcagacaacaaccagatcttaac  
 tttagaacagctgacacattggagccccctcccctctgccattgtcctgtaccttggca  
 actgactccagacctctatggagcttctcactcaggaggggacagagcgggtggttatagtc  
 ccaatatggtattagtagccgggcatgccaagtgtgcttgcagtttggggttattcaca  
 gatgactttctagaccattttccccaaccaagtgttgggtgtatcaacacttaaacagggt  
 50 gccatgggattatgcatttcagccttgcctgtcagaagctggctgccacagtatctggg  
 tggagttgcctcgtgggtcctcctcgtg

>scr\_sc-133556969\_1 (TOXMARKER Assignment: 31; SEQ ID NO: 31)  
 55 tgtacaatgggggataaaagtgtcaaagtgaatgagatgttgcctatagtttcttcttttgcg  
 tgatagagcaccctgacaaaaagcagcagcagaggaaatgtatctggcttacgattccat  
 gttaaagcccgctcattgatgaggtgggtcggggagtcagggtgaagactgtaaacagctag  
 tcaatcacatccacagtcagagacagaaggacacaaattcatggatacttgccttttgc  
 actcagctcagtttctccactcttacacagttttaaatgccttgcctaggagtgatgcc  
 acccagctggggtggatgttccacatcagttatgacaatctccacctcatgcccata  
 60 ggccaacccaatgtagacaatctctcattgagactctcttccaggccatgtcaagctga  
 cagttatagctagc

5 >scr\_sc-170142736\_1 (TOXMARKER Assignment: 32; SEQ ID NO: 32)  
agatctcttaagtgtgaaaatagaaaaatgattactaacgagaagatagacgcctacaacga  
agctgcagtcagcattctgaacagcagcaccaggacatccaagtcgaatgtcaagatggt  
cagtgtttccaaactcatcgcccaagaaacccatcatggagtcttgggtggcttacacct  
tcctgaatcaagcagagaaaactagtgcattgattctcatga

10 >scr\_sc-2563586\_2 (TOXMARKER Assignment: 33; SEQ ID NO: 33)  
tcatgactcccagcattgacattcccctacaatagggtcttgagccttcacaaaaccaag  
ggcctctcctgccattgttgctcaacaaggccatcctctgcttgatatgcctctcgagtc  
atgggtcattccatgtgtaattcttgggtggttagtacctggcagctctgcattggtgat  
attgttggttcttactatggagtgaacagcctgttctgcttggttcaattattgtctaaact  
ccttagttgagtaccctgttgcagtcgaatggttgggtgtcagaatctgcctctgtatt  
tgtcaggctctggcagaggtctctcaggagacagctatatctgggtcctttcagccagcac  
15 ttcttggcattagcaataatgtctaggtttagtactataaatgggatggatccctaggt  
gtgatagtttctggatggccttctccttcagtcactgctccacattaggtcttgatatttc  
ctccttattttgttcccttcttgcccatcggttgcccttttgatagattttgcagtt  
tagaaatacaatttacgtgcagggttattgcattcagatct

20 >scr\_sc-87618257\_1 (TOXMARKER Assignment: 34; SEQ ID NO: 34)  
tcatgatgaagaaatgggttctcggcaataggcaaaggcaggatgagagcagaggggtcc  
atgggggtcgaaggctgcccatgggggtggttctatgctctgaccattttgagatgaact  
aataatgttccggcagtggttatccctaacaaagatcacaaagccgctagtggagggaa  
tggaatctgaactctgtaccagcctccaagatccagatct

25 >cszr\_\_96561134\_83760493 (TOXMARKER Assignment: 35; SEQ ID NO: 35)  
gaattcactagaccagcatattgctctatgctgcctttccagcgtgtactgcctgtagt  
ggaacagactcttgaggtccacagtagagcttctgcacagcctcagcaaaaagtgttg  
tcacctggaaattggtgagcagagcaattccactgtccacagctgtcctccgaatcacat  
30 aattatcatggacaaaatttggtgttggtatttggggaggttaatcactaggtcaatgcttc  
cgtctcttatcaactttctgatggaagagaggtgggattctgtccttctcagatggcc  
aagccactgggtggcaggaacattgttggcgttgagccagctctgatgtggcttctgtgg  
caaaaagctt

35 >scr\_gb-x83855\_1 (TOXMARKER Assignment: 36; SEQ ID NO: 36)  
ctagtccccgcagcctagcgcgggcgggcgggcgatggaggagagcagagccccgggc  
cccgcgcgtcctccagcgcgtccgctgcaaccccgcagctgagcccagaggctccggccc  
tgtgcgcctaccgcggccccgccactatggccggcgtgtgggcgcggagcactcggtt  
gaagcgcacagcaaccagtcaagtgtgcccagcgtgcggtctgtgtccgtggccttc  
40 cccatcaccatgatggtcactggcttcgtgggcaacgcgtggccatgttgcttgtctcg  
ccagctatagacgcgggagagcaaacgcaaaaagtcttctcgtgtgcattggctgg  
ctggcgtcacccagcttgggtgggagcgtcctgaccagtcgggtggtcatcctcgtgtac  
ctgtcgcagcagcgtgggagcaactcgacccatcggggcgcctgtgcaccttcttcggg  
ctgaccatgacagtgttcggactgtcctcgtcttgggtggccagcgcctggccgtggag  
45 cgcgccttggtatccgtgcgcgcactggtatgccagccacatgaagactcgcgcacg  
cgcgcggtactgctgggtgtgtggtgtctgtgctcgccttcgcgctgctgctgtgctg  
ggcgtgggcccgtacagcgtgagtggtggccggcacgtggtgcttcacagcaccgggccc  
gcgggcaacgagacggactctgcgcgggagcgggcagcgtggcctttgcctccgcttc  
gctgtctaggttctggtgtctggtgggtgaccttgcctgcaacctggcgaccatcaaa  
50 gccctggtgtcccgtgcggggccaaagccgcgcctcgcagtcacagcgcctagtggggc  
cggatcaccacggagacggctatccagcttatggggatcatgtgtgactgtccgtctgc  
tggtgcgcgtattgataatgatgctgaaaatgatcttcaatcagatgtcagtagagcaa  
tgcaagacgcagatgggaaaggagaaggagtgaattccttccataatcgccgttcgcctg  
gcttcgctgaaccagatccttgatccctgggttctctgctgctaagaaagatccttctt  
55 tgaaggttctgcagatcagggaccacccaactatgctccagctctacctccttgcct  
tgcccagcttctcagtcctgtgatgtggagtgaccagctagaaagatgatgaacaacctga  
agcggagtcttattgcaatacctgcttccctgagtagagaatttcttccccaggggaag  
gataactgaatcattttggattgtatcttcttccgcctcatattttaagtttctctgc  
cattaaacacaccgagacaagctt

60 >cszr\_229800465\_190907286 (TOXMARKER Assignment: 37; SEQ ID NO: 37)

agatctctacaccgcaaaagggtctcttccgtgctgcgggtgccagcggtgcgtccactgg  
 catctacgaggccctagaactccgagacaatgataagaccgcttcatggggaagggtgt  
 ctcaaaggctgttgagcacatcaataaaactattgcacctgctctgggttagcaagaaact  
 5 gaatgttgtggagcaggagaagattgaccagctgatgatcgagatggacggcacagagaa  
 taaatctaagtttggcgacatgccatcctgggagtgctccctggctgtctgcaaggctgg  
 tgccgtggagaaggggtgccccctttaccgtcacattgcccacttggccggcaaccctga  
 agtcatcctgccggtcccagctttcaatgtgatcaacggcggttctcatgctggcgacaa  
 gttggccatgcaagagttcatga  
 10 >scr\_gb-bi277612\_1 (TOXMARKER Assignment: 38; SEQ ID NO: 38)  
 gggccccctcctgctcgtgctgctggaggcgtttcggcgatattacaactatatttttgg  
 tttctacaagagacatcatggccctgctaaatttcaagataaaccacagttagagaagct  
 tctggctttcattaacctcgaaccgcagtgatgccttccctagtatgtcatcagatga  
 15 gtccattctctacttgtacaagaaccagtagctctcctcaaggccaacgaagtttgggg  
 agcactaagaggtttggagaccttttagccagttggtttaccaggacgcttatgggacttt  
 taccatcaatgaatccactattgctgattctccaagattccctcatagaggaattcta  
 tgatacatccagacactacctgcctgtgaagacaatttttaaaactctggatgtcatggc  
 ttttaataagtttaacgtccttcaactggcacatagtggacgaccagtccttcccttatca  
 20 gagtataacttttctgagctaaagcaaggaagctatttcttctcatgtctatac  
 accaaacgacatcccatatggtaacttgaatatgcccggtccgagggattcgagtcatacc  
 agaattcgatagccccggccatacacagtccttgggggaaagggtcagaaaaaccttcta  
 tccatgtttcattcaaaaaattagaactcaaaagggttgacctgtagaccaagtctaaa  
 tacaacatacgtattctttgacacattcttcaagaaatcagcagggtgtttccagacca  
 25 gtttatccacttgggaggagatgaagtggaaatttgaatgttgggcatcaaaccacat  
 ccaaaatttcatgaagaaaaagggttggcaacaattttagaagactagaatcctttta  
 tatcaaaaagtaagtcactgaaagcctaataccactgttttcatacaagtccaagctg  
 cgacttagctctctgctttacttctcatcttccctactgcttgcaagagtggagccaaga  
 acaccttaggaggcagtaagcattttgcagtaactactgaaatagaggggagaagccatg  
 30 ccgctaggagctctggctgcccttctgtcttttgcactatccaggggctggaactcactc  
 cctttgtcctgagtgacctggggcatctctgctccttacacagtgcagtgacatttccaa  
 cattccacagccagggaattgggtactgaagtgggtgctgccttggttagaaaacacagaca  
 gaccacttcccaaaagtttgggtggacagtcgttctctaagaatcagcacatttttcccc  
 atagggaccagaccacacttaggcacatcatgggccatgtggagttgcaaatctcttttana  
 a  
 35 >scr\_gb-j05266\_3 (TOXMARKER Assignment: 39; SEQ ID NO: 39)  
 ttttttttttttccagagcagaggtcttttttaatacaatacaaaagtactttaaaatctc  
 ataggggacagccttgaatcatctateccacgctgattgtaccggtgaagtagaacaggata  
 40 agagcaattcgcagctgcagcacagtcgtgtacacgagcagcccggggagccatgcc  
 tggcggtacaatgtgctctcacaagaagtaactcatggaactcaacgtgaagtcgcgcttt  
 ttttttttgggtcttttttttccggagctggggaccgaaccagggccttgcgcttcc  
 aggaagcgctctaccactgagctaaatccccaaaccctgaagtcgagctttaaataata  
 acctgagttaaattccagggaaaggaggcactgactcctacaggctgctctctgacct  
 45 ccacaagtccaggaatacatctgagccggtccacacaaactagcactcaatatggaaact  
 tttattcatgtgatttctgtacatcaggagtagacaagagtaaacctttacaaatggtgct  
 gattttaccacaataaatgacaaaaccaaagcagtgctggtgacagtggcagggttta  
 aggttcaaaccagccaagaagtttggtacgatttcttcagctttgcatccgactgttc  
 50 tgagattttcccatcagacctgatattgcccaagaggctctggtgctggctcacaacatg  
 agacaagaaagcactctcgaactttgtgatcttactgggctccagtttatcaagataacc  
 ccggacgcctgcatagatgacagccacctgttcttcaatagccatgggagagtactgtcc  
 ttgcttttagcagctcggtcaggcgacgcccaggtcaagagctgctgagtggcagcatc  
 cagatcagaaccaaactgggcaaaagcagcagcctcccggtactgggccaactccagctt  
 catggtgctgccacctgcttcatggctctggtctgggaggcagatccgacacgggacac  
 55 agacaagcccacattaatggcaggcggtatgcctttatagaacaattctgtttccaagaa  
 gatctgtccatcggtgatggaaataacgtttgttggaatgtaggcgacacatcaccagc  
 ctgtgtttcaatgactggtaaggcagtcacagagccaccaccaaaggaatcgttcatctt  
 ggctgctctctccagcaggcgagagtgtaggtaaaacacatcaccgggtaggcctctcg  
 accgggggtcgggcgagcagagacatctggcggtgaagcaacagcctgcttgataa  
 60 gtcgtcatagatgatcagagcgtgcttccattatctcggaataactctccatggagca  
 gccggagtaaggagccaagtgaagcggggcagcatcagaggcagtggtgacaccac  
 aatggtgtacttcatggcatctgcgtctgtcagtccttccacaaactgagcaacgggtgga



ccgtttctgaccaatagcaacgtagatgcagtagtcttcttctcgtcagtcctccatc  
 attgaaacgcttctggttgatgattgtgtcaatagcaatcgagggttttcccggtctgtct  
 gtctccaataatcagctcacgctgacctcgccaatcggcaccaggctatccacagcctt  
 gatgccgctctgcattggttcccgacagagattcgggggataattccaggggctttcag  
 5 gccactcgtctgcgaatcttggaaaccaactggacccttcccatcaatggcatttcccag  
 ggcataactacacggcccaacagttcatcgccaactggaacgtccacgatggctcctgt  
 tctcttcacgatataccttctttaaattagcttgtcattcccaaacacgacaactccaac  
 attgtcgggttccaagttcagggacataccctttaagccggaagaaaactctaccatctc  
 ctacagcttgaacgttctcagtcacatgcactcgggcaataccatcaccaatgcttaagac  
 10 acggccagctctctcaaggtcaacagaagtacagctccaaggatccgctcctcgagaat  
 ggaggacatctcggcagtgccagctcttctgaagtcagtggttagaggcatggagatttct  
 tgtacaacaaaaagatgaccccaaggcatttttggagaccagtcgccgctcgagggag  
 ggcacggcg  
 15 >scr\_gb-m37394\_5 (TOXMARKER Assignment: 40; SEQ ID NO: 40)  
 ttttttttttttttttctgtgtgtgtgtgtgtgttacttcatgaaatgaaaacagga  
 aagcatattaaaactcaaaacaatgaaacagaaaacataaaaaggtagtctaatagtcaga  
 aaacactggtaaactagcgtgtgttaagtatcagggacatatttatacaaaaaagtaagt  
 20 ctgagggaaaaattctaccagtcatttcttctccagtcaccagtaagtaacaaagtggctt  
 atcctattgtaacctgccatgggtttaaagctgtacaagtgtggcctgctgagcacatccag  
 gacttcttgtgcatgtagtcttctgccatggaagtgtcttgatgcagagctgctagaac  
 caactgtctggtcagttggctccaggcaactctgtgtaatacacgctacgggcaagcttc  
 ttcttttatggaagagtgcataaatcaataaaagacaagaatcccagagttcccta  
 25 tgcagcaagcgccataggtctgtttttttccctctatgtacctcaccatgaggcaacc  
 ttctgttccaaaaggacaatgttctcgatggatacctttcagtggaatcttcacagttcg  
 aagaccaatagataaccttcaacttcccaaagagcatcaggggagggggccacttcttg  
 gctcagtgacaaagcccgtcagagttatgtctttaaagccagtcctgaggggttgacatttg  
 acacaatgtggacatggctgtcaggagcagaggtgtgccatggcttggctcctgggctc  
 30 tggaaagtccggttgttaactgggtacaatgcctcttcaatgtcatgtctccactaaactca  
 ctgcttggcggtgccaccgcaggtactctgcattttcagctgtggggcccttaaagatg  
 ccattcggcttggcttcttgggaaagaagtctgtgtgtagtcaggggtgtccaggctc  
 atttgggtggctgccttctggatccagagggcagagctgtcaaaccactactgaggcag  
 gtcggctgggcagtggtgagatactcaggggtgtctcaccgcattgtatggggattttga  
 35 taatgcaggtctcttccaggagctggatgcaggggctgattgtgtagactgggttctgc  
 acagagccagccggcctcttgggaacagattgggttatataattcaggcacgggaaggaa  
 gtgtcatctatgttgcctctgtcaggacgctgggtgggatcggagctataccgttgcaag  
 aaggcgtcttcttggacacggcagctccatttctattaatgcaagccacagtggaactg  
 ttgctatttgcactcagagagctcaagagtgaggtccgtgacgtggatgggctgttgaag  
 40 aagccttgcctgtgggatgaggtattcatcagcatcaactacgtcttccatgtcctcctcc  
 tccatcagggctcggtaaaagtggagctctgtagggctcggcaaatgcataccttcatcc  
 ccttggaataacaaggtagcgtgtgggtctctggccattttggagaattcgagaatcaac  
 tctcggaactttgggtggctatcagcatctatcatccagcacttgaccatgatcatgtag  
 45 acgtcagatgggtgcagataggtggctgtggaaggcgtctccttctctaggatggatgag  
 atctcacttgcaggatcccatcataaggettgacccaaaggctcatcagttccacacg  
 gtgactccatagctccagacgtcgttgggtgtgtataaattcgggtgtaaaattgattcc  
 aaagccatccacttgataggcactttgccccctctgcattggtatttcttctcctcagca  
 ccaagcagtttggccagtcacaaatctgtgatcttgacatgctgtgggtgtcttaccagt  
 50 acattcctggctgccaaagtacgggtgtacaaacgcccgtcttccaggtagttcatgccc  
 tttgcaatctgcacacaccagttgagtaggtactgggagccaatgttgtccttatgttct  
 cggacatagtcaggaggcaaccatagggcagtgattgtgtaatgagctggacagtgag  
 gtcagacagatgccaggaggcggcatacatgagggtgtccacactggccatcacgtag  
 gcttcatcaaggatttcttctgttggcttgggagatgtggcttctcttaactccttgatg  
 55 gccacagggttttctcgtccttctgggatccagagacccttatacactgtgcca  
 aatgctcctgaaccacagaactttgatcttttgaattctgttctctttaaatacctcaag  
 tgggttgggtcggagcttctccgctgggtgtgagaggtccacagagctctctctctga  
 agcaggcggcgtagtgtagcttttcggacaagctgacgtcgacgcataagaggccgatc  
 ccaagggccaccactactatgaagaggaggccaccacaatcccagtgggcatggatggg  
 60 atcttggcccttctggttgttgacatctttaaaggcctggccagcacatccataggta  
 cagtttgcattggcagaggtggcagagcttattggcatctgcaaaactccagaccagggtg  
 ttgttctcccccatgatgccgaagggcaggtcttgacacagtggggaccatcaacatag  
 tgggcacacttgatgcagttgtctggccccggcctgtacaggtgatgttcatggtctgg



cttccacaggaaccgcccgcgcgcgcgtgtagccggtggcggcgatcggaacgagga  
gggaggacaacacagaagagagaactaaccgcccgtagaaccacgg

>scr\_gb-bi294409\_1 (TOXMARKER Assignment: 42; SEQ ID NO: 42)

gtgcacagaggggactcaacgggtgtgccgctgctcagactacatctggcccacaaatgtt  
cttctagagccaccagaattttaagattattggctttaaggaccacataaatgtgatgatg  
gagtttccacctgccacttacaagctattcggggaaagcttatggaaaagactggagctc  
acatccttcgtcatcgaggaacagacagaggacagcattagggtgcacaagccccaaatg  
aataatgtcactgggaacttcacgtatgtccttagagacttacttccaaagacaaactac  
tgtgtgtctgtttatttttgatgatacacctgtaataaaatctcccttaaaatgcaccgtc  
cttcagcctgaccaggaatcaggtatggctaggcttttaaaatttgcactgttgttttga  
tggaaaacttgcgtgaaagaaaaaaaaaactcaagttctggtacactaaatgtacttctt  
ccaataaatgcacatcactgagctgtttaaaaaaaaaaaaaaaaaaaaa

>scr\_gb-ab015747\_3 (TOXMARKER Assignment: 43; SEQ ID NO: 43)

tttttttttttttttttgacaagataaagagtcctttattgacatagagctccacgtgacc  
tcttctgtcctgccctccttgcaaacatactagggtgtccaaaggtagggacacgagcag  
acagtcctgagcctgggtcccgtcctccagaatgcagtcagactgcagtcctgccatctgcc  
atccctatcatctggccaccaaccagaaccagccccacagttcccttgtgggtctcgctt  
ggctggcagtggtgtgtccactgggacctggcactaggctgctgtgtttgtttactggg  
atcccatctccacatcctgggagccctggcttctggccacatgtgggtaactggcagtg  
ctttgggcaatcaagtttgcgttcttgttgccttccacaactgggccaagctgggacagc  
aggctctgcttctagtctcagtcagagctgttcaatgaatagcctccttggggcagtatc  
taccctcccttaactcaaaatttccactagttagggcctccaagccactgccaggccag  
ctgcgagtttctaggaccagcttccagctggagaaccgacagctatgccaggactgctg  
tgagccttgggcaaacgggtctattgggtggacagaatgggcctgagcaggtagggaaca  
agagctaggagagcccagggttaagaatatcagcactgctgtgggagaaagcaaaatga  
gtccctgaatcccttgtgaggaggagagcccaggccaacggtaggggagacagccaggc  
tctgaacttctagggtcaggccaagttcacatcttcacttcaccattcttctcgattctg  
ggaaacctgccagctgggtgtctctcaggaagcacttccctggcttggaggaaacccgg  
ccttagcacagacctcagcaacaacagcacactcacctaagacacagtgcgcccagagt  
gcccacaggtacctcagtagtctggctgggaacaggagagtggccaggggccttggccac  
ccctgacaaattggagggtgtcctgggtgctaaggtagggttggcttccctgtgacatttc  
cccaggacagctctccaaggtccccgagagattccccaaggatgggtgatttttcatcata  
gcaacagccgcagccagggttagcaacgacatggatctgaccatcttccctcctggctgtg  
gtgttgcctcaggtggccacacagatggcaggtgagggacaggtacacaatgccaaggcc  
caggtcatccccaaagggttgggcacctgctctgaaggaggaggggccttcagccaggt  
gctcctttgagccccagctctagaccaaggcattctggggtgctattgggtggggctga  
gttcaggggactgggtggcagctccatgtccagtcggaaggatgaataagggtgaggtt  
gggggactgggttaggaacagggttctggaagggttgtaccctccacatccactgtcagt  
ccctgctgctgctgtgtccgtgcagacgccactgctgctgagcaggctcgagaaagcctt  
gtaaccagtgtctccagaaggcccagaccaggcaccacctggcctgggacgcaccctg  
cttactgcctgcacaaactcttggtagccactggtaggggctgggggtggagccagctgt  
cccgtgctgcaggacactcatgtgaaggatctgctcccagctctccgcttgcctgattgg  
tggccctgaagaatgggggtcaaccgggtcagaagatcccttcttccagatgtccagc  
ctgcttctgttctgaagccagctctccaggatttggggcggggctggagaagtactaaa  
actccggtaggcaggattgtctgaaatgacaagggggacctgtgtgcaggctgtgccagt  
tgctctctcagggtctgggtgtggaggctgctgcctgtgacctggcatgtggtctcact  
gggccccgtggggaagcaggcccaggacgtagaagcttgccacttctgaaggcagaag  
ggaggatgactctgccatgctcgactggccaacgcctccattctcagccccagcaagtc  
tgaaaacaggttctcagtgagccgggcatgatgtctgcctgactctcctggaagcccc  
tccgctgttctcaggtgacatgctcaggtccctttgaccatctcatcctcttccctcctc  
cacattctgtactggggcctcaaacagctccatacagcgcaccacactgacatgaacgtt  
ctctggccagaggacgggtcctgctgacctccgcaggataccagcctgcttttccaggact  
ctggagaggcttgggttttggcagccttcgggattctctctcttcttcttactctatgctc  
cagcaagcaggggcagcagcttgggttagacaagtcttccagtgcgggtcttgggtgactc  
ctggcttcgggtctgcttctccagagggacaccttcgtgtcctgaatgatgatggctgc  
taagggactgcgtgctggagtgggaatctgggtcccaccatatcttcttaattcttgataat  
gctgaagtaacaggtcaggcaaaacaataggatgcagatgcaggagatgctgaccccag  
cgggaggcgctgcagcaggggcagctggaagtgtgtaccacgtgatgctgggactcca  
ctcactccagatgccagggaagctctgggacaagaccctcacacgtgccttatagcgac

cctgatgttaggggtgttgactgggaagctcagcttgggttccgtgtaggtcacattata  
 gactttgaattccgcgggttgtcctctctggagatgttgaccatgcagatgaggccttt  
 gtgcaggaagttgttcgatgggtatgggttgcctcacatcagcagcaggccattggagac  
 attggtgtggagtgtaggttgtctggagctgggggttcacattgtcactaggcttgaa  
 5 ggagccttgccacagctgtccccgctcagaccacagttccagccagtatgtgtctgcctg  
 gatcggtcctctatggccatttggcacacacacaggtgtcggcactgttcttgggggt  
 gcatgtgaggttttcagagaactcgaagagcagcctgtagtccaggaggagctgagaact  
 gcagtcacagtgctatccagctgccactcacacgtagaagtgcggatgtagtgcagagaa  
 10 gcaggtgggggtcaccaggaccttgatgccccagagccagtcacccatagcaaaatcag  
 acagctcacggaggacaggaacttgggtgcaaagccgccccattgcggaacaaaaggtgcc  
 tgggctatacaggagagactggaatgcagctcagtggcagcgtacctggccccagatc  
 ctgggctcctctccagcacctgtgtgttcaggtccacgcgcgtgcggggctttcctg  
 cgcgaaggacctcgcccggtttcctacgcgcgcgggacgc

15 >scr\_sc-191879433\_1 (TOXMARKER Assignment: 44; SEQ ID NO: 44)  
 gtgcactaagaatgacaaacttgctgtgtgccacaaagatcttgggtggctgggtgggtg  
 ccagtggtcaggttggcctcacactgctccaagtagaagagcagcagctgtcggctgaa  
 ggccccagtcacctgtccgccccggcacaaaggggtgggtgtccagttggccagg  
 20 tcatgggtctatgggacgagacacctcctgctccagtcgctcaaaactgtttcagctgtgc  
 agtccagttggccttttccctgtgcgcagatgttgcccttttccagcagttccttctgg  
 gtcttctcaaattcctccttccctgcagatgaacgtagtcatagtctcctccatccaaccc  
 ccttcactgttctcactactggccatccgga

25 >scr\_sc-140438096\_1 (TOXMARKER Assignment: 45; SEQ ID NO: 45)  
 tctagactttaacaacaagcgtgatgaacacccagagaaatgcaggagtcggactaagaa  
 catgatgtggtacgggtgtccttgggaccaaagaactgcttcacagaacctacaggaacct  
 ggaacaaaaggtcctgctggagtgtgatgggcgcccagattccccctcccaagtcttcaggg  
 aattgctgtcctcaacattcccagctatgctggagggaaccaactctcgggggggaccaa  
 30 ggaagatgatacttttgcagctccatcattcgatgataagattctggagggtggctgtg  
 gttcggcagcatgcagatggctgtgtctcgtgtaattaagctacaacatcatcgaattgc  
 ccagtgtgcacagtgaagatct

35 >scr\_gb-x87157\_5 (TOXMARKER Assignment: 46; SEQ ID NO: 46)  
 tttttttttttttttttttaaccaagaggaggaatataattgtgataggaaactaagaatc  
 atgaagctcactacaaaagacaaacactactgaaacatgttgtgctggccttgacacacg  
 caggcagactgtcgcctagctctgaggcagagggtcaaggttgacacagggctcggagga  
 aatatttaccagagagaatgtggtgattcatttatcagtcacagagatcgcaagtataaaa  
 cttcaagatataagaaggatcaaattatatcatgtatgtgattcaattttaaagtctta  
 40 gccctcttacattatattatctggattataactgtaaaaaaatcaaattacattcattat  
 gaaactttttatcaaaaagaatcaaattcatttttatgaaaactttatagtacaattat  
 tagttgggtctttccttaggtcacagtatttataattccatttacatctgtataattttta  
 aaattaaaaaaacaaaagcaaatcaatagaaatctaagttttcttttgtaaaactctcttc  
 agtctccaggccggcaccacatgacagtgttgacttgcctccagacatggacaactccc  
 aggatccctggcttacgaaccattcaggcctcgactcattaggaatgctttttgggttgg  
 45 ctacagttgcaagaattctggagcatgtccatgccgtccagggaacccccaggccttcag  
 gattaagtttctgtatttcatccaacctctggattcatgatccccctcttttttaaaaca  
 gctgtgaaacatgtccatggaaaacacttcactccaaagatatccataatattggccatc  
 ataccctcctgccaagtgtccaaaagtagctggcatatttgtgcctggcgtagctgcaac  
 tcccagaatttctgtgcagtttttagcgtattcgcctcgcgcatccagagtcgcattgggt  
 50 atggagagattgggtcaactttgtcctcctcctcctcctcctcctcctcctcctcctcctc  
 gaccagcctagaagcaacaagcttctcgagcagctcgtctgtgatagggtgtccatcttt  
 ataatgctttgacagttttcgcagggaatcaacgtcccacaccagttttcaagcatttg  
 tgatggcacctctacaaagtcagtttccacgtttgttccactgaatcgtgcaaagtcagt  
 ctgocgcacagatctgatgcatgacgtgaccgaactcgtggaagtaagtcgcgacttcac  
 55 atgtctcaggagagagggccgacctgctacaggctgagagaagttgaccaccagggcggc  
 cacagacatcatccgactgccatcagggaagagcagcctggctggagaccgaagcaggc  
 tgcattggtgtattttccttcccttggatagaggtccaggtagaactgccccaggacctc  
 tctgtagctttatccttccacagtgtaaagtgaaacgctcttattccaaacatgagcatc  
 gggcacttggtccaaatgaaagtcacagcgtcctggtagatgcttagcaagccttccgt  
 60 gaccacctcaatggggaagtaactccttaagggaactcctgggtccaccgagtaacttgagctc

ctctgtctgtgcatgtagtaatggaggtcccatgcattgatcttcccgctcgtattcaaa  
 acctcgctctttcacattccttcttcttcagggtcaaaaataaactcccgttctgcctcacc  
 caagggtttcaatttctgggttaaatacatctagaaaggcgccacgcgggtggtgctctt  
 5 cgcagtggtcagttcaaggacaaagtcagcatgggtgttatagcccagcagcttggccac  
 ttgagctcgcagcgggaggagctgttgagaattgcgggtgttttctgtttgcacctgggt  
 atgaaaagccatttccatcttctctcgagtttcagggaacacagcatttcttcatgacagg  
 gaagtagtgaggatactttaaggtaactttgtacttgtcttcatctgttttttctaaact  
 gtcaatgaagtcacaggaagagcaccaagttcagccttggagaatacaagggaagtgtc  
 10 gtccctcattgagggttcttgttgaaagtcatactcactcatttcttcttcttattga  
 tttgatttcatcttcttattgtgttctgaaagatggagttccatttcttcttcttcttctt  
 tgacttttccaagtatcgcttgggttcagggtttatcttcttccaaatcgcatgtttcttg  
 taaatgaacaattctctgaaacacatcttctctcatgtctcatctcaatatcaaaacgaga  
 aagctttttgtctgttctgtgcttgcagcccgcaacttctctgtcagatgacacgtgctg  
 15 agggaaagtcacagcatgggtcctttccactatgtacgtcacttctatgtcagccagcacctg  
 cagacagttctcataagttacttcttccagggtgattgtccccacgggtgtcgtacacctg  
 cttgggtctgtgctatgagctgctctgtcctcgtcttgatctgctctggagaaagggtccca  
 tctgagaacatttctgcccagccgagtgtaggaagacatagcttgaagaggagaagccag  
 ctcttcttccagtgctcattgtcagctgaagcctggagccaccagctctgtggaggcctcg  
 20 cagagtcgaaaggcacagggtgatcatgggcacgcgggaggccggcagcagctggcgcg  
 tctcctcccgtttaggtgcaggaggcaggcggtggtgtctgcccggcccggaagccag  
 gagtgggccaagccgaggagaccagatctcgagacggaggccgctcagtc

>scr\_gb-u66707\_2 (TOXMARKER Assignment: 47; SEQ ID NO: 47)  
 tttttttttttttgtttatatgccaaacataaccttgtgctagaaatactttatgggggtt  
 25 acaactctttatatacaattttttttgaggcagtatctctgatggagagcataacttgta  
 aagagcttgtgtgtgcttccgtgctccaaaatgataggaaatccactttgagaagacaac  
 ttatttgattttaaaaaaacaaaaacaaaaacagaaacaaaaccgcaccaatgc  
 acagccagaggctccgctggaactgatacagaaccgcgcaaacgcctgattataagtaa  
 30 cattttccagggtgggtcaaggctaactacataattatacacctggcactgatgtttgcc  
 attgggtcagcaactggcaaaatttgtttctatgtataaaatttatttttaaacattatctc  
 tggcctgacatacttccactatttataaaaaacatttagacagtgagctcacgttgaataa  
 ctaggctctactgtgttctggaagctcttcagtagtaaaacagctttttcgtgttccatat  
 gcacaaaactgtgtccatttgccctgaaggattttatccccgggctgtagaagggttgatg  
 35 ctgggtccatcagggtgaaccctagtaacaaagataacccttgtcggaaagggttgaaaggat  
 ttccttgcccactaattccaccgctgataactaaatccaagcccagggttcttttctattc  
 tcacacagaactgctcgggataaccgctccatactcctctgccccttctgttgaattaaagc  
 accgtccagggtgggggtccccgggtggcctgtgatgaggggatctggatgggagtgccg  
 actggaactgctgaatgggtcactttgttgatgttcccttcatatggctgctgctcccggc  
 40 tggcgtgctgaaggctctgtgaccccatcagggtctgaatgtggctgcctgccttttttag  
 taatgggtgtccggagggtacatcccgctcccaagtgggtgaaggattccattggccactag  
 gagagacatcttcttgtccattgtctaaaatgttgctttgctgggacggggtcctgtcta  
 accttctagcttctatattgtctaagcagctgctgtctccagctctgcccggcattttgccac  
 agctctcctctcccttcacaggggtgggccttgtctttatatcactgttatctgatgtct  
 45 tgtcaccatagttacccaagttatagtcagatgggtattttttccaggagggtgcatgg  
 taggcctagctgaaactggcctgggtttgggaggccccgtaactctctgtgctgtagctcc  
 tggcagacagtgggcctcctctgggtaagggttttttagcaggaaaacttccgcctgctt  
 tcacttcttgatatgaagggtgctcatcttctgacctgccattcctcttgaggaattggg  
 gatcagtcattgaaacgctgctctggccttccagccctcccttataggctgctcggccat  
 50 acctgtcaccagggggcagctcgtggggctcactgaccttctgaacatggccatctctg  
 tggagctggccagggtgagcctccttaggaagcctgccttgggcccctggctggcga  
 actgagcattcaccatggcatcctcatttacagatgggtgggaaaaggagaacatctgct  
 ccatgggtgggtatcctctataggctctggggtgacgagatccttggcgatgtttttac  
 cgggctgtttgtacatactcagaattgtgtgcaaaaggggtgggacctcttctctgctt  
 55 ttacttccaccgctccttgtggattgaagctttgggtcaaactgatagaccttcttagtca  
 tagatgctttttgttgggggtgggcctttactacttccgtacatgagcatttctgctacca  
 gcatgggtaccgactggctcctggacatactggacatgccgtgctctgggtcccaagaact  
 tgtctggccgctcgtggcttccctaagtgtatcgcttccagaggcatagttttccagtgga  
 60 tgttatagacctttaggtaccgacgtcaatctcatcaatactctgggactttttgaact  
 tgttggacttttatatctttcatgagcggggaaagcctctccgtgcttttgctaattgcaa  
 taacaccttttagacgaatctgggcccagtggtttgagaaaagacattaccgagactcc  
 ggttgggggtgggggtcgtggtactcccatggcactcccgagaaaagggacctgggtgctt

ctgtaggctccttcatgtgggtcttttctttcaggcaagggactggtagtaggggttgttt  
 ctaatttggagggaaaagcagtcctgtcctcaaaggactgggggttctgggtccaattct  
 gccagggttggaaggaggcacttctgtttctgggtgtgtgtctgtgggtggactgctcca  
 gttccagggggaacaccgacaatcctttcttgctgattaaaggcctgcgcccatgagcag  
 gtacgcttctagccttggagcttaagagaggggttattgttggcattctcgctgtggctt  
 cctcggagacaaaacctgtgttatcgtaatgggagccatcgggtccagttgtcagggaaag  
 catcactcattggcagccgatcaggacgctgggggaggttccctggagggacagcctccc  
 gttgggtgagtaatggctttgcatctagaggctgtgggaaagggtggtgcaatcctgttac  
 ccacaaaagagtattgcacggcatcttttagttgttggctgcaaggaccccactttcaccc  
 ggggtgttggaggatgctgaggaagcctgggagggcgagtagtctgagtaggtgcctgagg  
 agacactgttattcagacagttagttttgtcaacttcagactcatcagttgattcttttt  
 tgtccttccctagcagaacaagtttgggtgggtacagaggggtctcagctaataagggtt  
 gaagttccccaatcctcatttcttagctggatgaacaaaagaatcctccactgtaatct  
 cctttggggccaccggccacttgtgttcaaacttttctttcacagtttgctccgtgttag  
 ctgttgggtttgcttctcaacgcgcactccatggcttggcttaccaccagattttgaa  
 cagattttaccatgttcttttaaatcctccgggttaaggagttggatatcgttttaggttta  
 tttcaacctgtctgccacttagagagggaagagtgggtggattgggtgccactggtaatg  
 gagcacacatgctcctctcctgctggaggccacttatacaaccccatgccagctgggggt  
 cactctgggggacgggcataatcttggatctgctgatcacacctggcccatgggtgtgcagc  
 aatcgccagacagcctggcaggttggagagtaatcccacgctggccctgtcccaggggg  
 cttggcaggagagagccttaactttccagcactttcgtcatcttcttttttatcctcaa  
 attcaaaggcaacagtcagtcgctgttgtctctgctcctccacaggggtgggggtgaagc  
 tgtcgctgtctgactggaaatcttcatcaccacggggctgctggggaaacatgtagtgg  
 tcagtaccctttgcttgggttctggatgggcttctgtttgcagagggatgagggccttgg  
 actgattgtcagaaagccacaatgctgcaagctctttgagtttgggtgaaggagaatggca  
 agttcttcaacctattatcacttagatttaagactcgaagtcctctgcatctgccgattt  
 cttcaggaagaaattctagcttattggagcgtagagacataacgggtgacgttcttacagc  
 ttccaatttctctgggcaactctgggaggaaattctcgtccacagctaagggttcgcaggc  
 tgtgcaggttaaccaatgggtgggagggagggactccagctcattgcaactgcagtcgaatt  
 cttctaataaagataagtttccgattgtgttgggtagcattgtaagctgattgtcatcta  
 ctttttagagttgttaactttttcagcaatcctatagagtcggcgagctgttgcaacatat  
 tggatgatagtaagaggtcctcgagggttcacatccagaaatatccatgtcaaccgttt  
 ctatcctgttttttgacatatccaggtataccaacatctttaacttccctatagaccag  
 gcagcacttgcaatgcgttgttatccatccacagctccctcaaattctgaatttgatcca  
 gaacttcaggcagctcgctgaattcattattgcctagggtcaagtccttccagctgggcca  
 gcttgtgcatgtgactttggtagagtttcaagtgtattttctcttaactccaagattcgca  
 atttgacaagtcctccaaaattagctggaagaaattcgaggaaggcgtcattcaggtaga  
 gctgggtcaggttaagaagctgcgtgaagccatcgggtagtttagaaatgggattgacac  
 tggcttcaataatgggttaaacacttacagcactttatgttttctggaaattcttgtacac  
 cgtttttactgatgtcgagttctttcagattaaactaggctagcaatggaggtcggcagac  
 ttgagaggtcattatcaggaatgcttagtttcttagagcttgacagttgaacaattgct  
 tgggtagctcctcaatctgattggcatctagatagagctcttctagtgtacgttcgaagt  
 tgaagacctccttgggtacctgttgcaggctgcagtgaggagtaatccaacaccgagatga  
 tctcttctcgccacggaagcagcggcatggcaccaggcggcggatgagcttccgtttgg  
 tggctactctccaggcactgcttagtctagtcactcctgtctctgaagacttctaggctgtg  
 ggcactttgacttgcattcttttcatgtagcgggctcactcttcttcaggcctcttccga  
 agtgcgtgcacgggcctccttacaaggacttctctgatattgtgggggattccttccccgt  
 attaggttctccatcatcgcaagca

50 >scr\_gb-af017393\_2 (TOXMARKER Assignment: 48; SEQ ID NO: 48)  
 ttttttttttttttttttttactagtaaggtatttactaggaaatgatacaaacagccag  
 gaaaagggtgcatcgcagaacagggtctgtgcgtataagatgggtatttcccccttgtca  
 cgtcatttttttccatgaagatgcgcttaagataggaagggtaaagtaccgacacgtggca  
 gggccccgggttttagggaagggaacgtgagagagcgtcaatggaggccacacagtgaa  
 55 acccctggaagagggccagagcagtccttgggtgcagtactcagcgaatgcgcatacaca  
 actggaaaggccttggcaaatgcccagccctgagctgagcgggggtcaggtcgatgtcct  
 caggctccaccagcggatgcaacgtgaagtctggagaatggaggtgaggtatatgaaca  
 gctccatgcgtgccagtggtctctccagacacagtcggcgctcccgccgaaaatggcatga  
 aggcgggggtcttcttgaaggattgattggcatccagaaaatgctcaggattgaactcct  
 60 gaggggtcttgaattgggtcggagtcagtgacaggtgttaaggagcgtgatgacatctg  
 tgccttgggtatcaggaagcccctgaaaggtgtgtcccgaatgacgcgggtggggcaggt

tcatggggatgacgtctgcaaagcgctgcacttcgtggatcaccgcgtctgtgtaaggca  
 tggatgcacgggtcctccagcgtgggcatccgcgaacgtcccaccacacaatcaatctctt  
 cctgcacacgggcttgcaactttgggggtacttcataagaatgaggaaggcatggcgtaaag  
 5 tggtgccacagctctccgttccaccaaagagcaggttgtgtgtggtcatcagnagggtgt  
 ccatattgaagtgggtcagtggttctgtctcctgtaccatttttgtgaggaagcagt  
 cgatgaagtcccggggagagttgggggtccaggagtcctggtgctcgcgagcgtgcggg  
 cgatgagatctttcatgcccccaagttccggaacacgcgtctgtgcggccaggcaccc  
 agtccaggagactcgggaagatgttgatcatctcgccccagggtgctcataatctgga  
 10 agttgtcattgataaagtggataatggtgagcagccgttcacgtcataatcgaagcgac  
 tgccgaagatgacagagcaaataatggtggagaccgagcgggtcaggataaacacggggt  
 caaagggttgccttccgttttccgcagcagctccagcaggaagctgccttcttccagga  
 tccgctcctcgatgcttcttttcccatgccaaagttcctcaggatttggaagagaacc  
 ttcggaggatcttccagcgttctccatcggaagaaggcgatgccgttgccttgggtgaagt  
 15 tgaaaaagatggggtatgagcctcgccactgaactcctcccccttgtccacaagagcct  
 ccttcacagtttgatatccgctgaggacaatcacacgcctgggccccagggtacacgtga  
 aactgaccatagtccttgtaagcttgggtgagtgaggtcagcaagcttggggagcgaa  
 gctgcagcaggtttcctaggttgggagaggcttgggtcctggaggagctggccttgc  
 cccatgaggtgaaggtcaggagacagagatgacagccaggaggagaagcaagatggctg  
 20 tgctcacaccatccatagtgaaggcagc  
  
 >scr\_sc-134241980\_1 (TOXMARKER Assignment: 49; SEQ ID NO: 49)  
 actatatgatcctgtttacatgaaccatacatactaggcaaactgtagacatagaatc  
 agaccttatacatagtccaatagcatagatcacagagcatggagacctgataaatgggga  
 25 ctgaggctgttgggaagaagtgaggaatgactcagcaaccttgggcctggtctccagcag  
 gtctcccagaatcagaaaaatggggccattttgaacagaagtgagtcggctgactgcctc  
 agcacaatcagcgggtacaaagcaaatcttgtaactgagtcacaagcaacactctct  
 gctatggattcctgctcatgctcaagtaccctcatgttgcagagaaagtccaaaaggaga  
 ttgatca  
  
 30 >scr\_sc-191609675\_1 (TOXMARKER Assignment: 50; SEQ ID NO: 50)  
 gccggctcaaaggtctctgcgagcgcattgggtgttttcaatgacaatcttgcgtgccaa  
 g  
 tcttctcccaaaaaggcgaattcatccagcatttcattgggtggttctgaaatgcgcttt  
 35 ggcagtggcgctgggtgggcatcttctccgtgccaatgggtccggttgatcatagccct  
 tgaccgagactacggacaatgatctcccgatagatct  
  
 40 >scr\_gb-x17037\_2 (TOXMARKER Assignment: 51; SEQ ID NO: 51)  
 gaacacagacaaggatgtatgtgtgggttcagcagcccacagcatttctgctcctgggac  
 tctcacttggagttacagtgaagctcaactgtgttaaagatacctacccagtggtcaca  
 agtgcgtgctgtagtgccagccagccatgggtatgggtgagccgctgtgatcacaccagg  
 45 acactgtatgtcatccatgtgagcctggcttctacaatgaggctgtcaattacgacacct  
 gcaagcagtgtagacagtgcaaccaccgaagtggaagtgaactcaagcagaactgcacac  
 ctactgaggatactgtctgccagtgtagaccaggcaccacaaccccggaagacagcagcc  
 acaagcttggagttgactgtgttccctgccccctggccacttttctccaggcagcaacc  
 50 aagcctgcaagccctggaccaattgtaccttatctggaaagcagatccgccaccagcca  
 gtaacagcttggacacagtctgtgaagacagaagcctcctggccacactgctctgggaga  
 cccagcgcactacattcaggccaaccactgtcccgtccaccacagtctggcccaggactt  
 ctcagttgccctctacaccacacttgggtggctcctgagggccctgcatttgctgttatcc  
 taggcctaggcctgggcttgcctggctccctgactgtcctgctagccttgtaacctgtcc  
 55 gaaaggcttggagatcgcccaacactcccaaaccttgttggggaaacagcttcaggaccc  
 ctatccaggaggagcagaccgacacacactttactctagccaagatctgagcaataccac  
 aggagtggattttatggggcacagacagcccatatcctgatgcctgcctgccagggccct  
 ccacaccgttctaggcgtgggctgggtgtgcactctccatgtatgctgtgcatactac  
 ctgcctgggtggcactcctaataaacatgctcgagctgtgagctgtcactggccctaaa  
 aaaaaaaaaaaaaa  
  
 60 >scr\_gb-bi291805\_1 (TOXMARKER Assignment: 52; SEQ ID NO: 52)  
 ttttttttttttttttttccggggtcaagatatctcgtatgctttcagggttgaattc  
 aggggtcagcaagggggaggggcagggaagggaacacacagggcatttccaatcactgt  
 gacttctggcaggtctcgatgcttctcattggccagtggtgactgatcagttgggacatggg  
 gagaagtccgtgccccccacgtctccattgaaatcttcttctgatatttatgcacatca

5      ttgctccggtccccgtcaaagtttccacaggccccacacaacatggcgcataatgctca  
tcaaccatcacattcagatgcccattcctttccaagccacacctggactccggccttctgg  
tggacaaacatggatccgtctgagatcttctcacagacacagatgttaacacagtagct  
gggagatccaactcggagaccattcacccatgcacccttgcttgggatcacagtcacccat  
gccatcctggaagaagatgtggaccttgctcacgatcttgctcattggt

>scr\_gb-aj000696\_5 (TOXMARKER Assignment: 53; SEQ ID NO: 53)  
10      ttttttttttttttttttgggttgggttgggttgggttgggagacagggtttctctgtgtag  
tcctggctaccctggaactaactctgtagatgagactggcctctgactcaagagatctgc  
ctatttctgtgaggattcaaagtgttcatcgcaatgcccggcttagaaaatgagtcttga  
aatggcactcagaaggggtggatgtggccttttgaacgggcaagtaacacaggtaaaatga  
aaacacaacagggtgcagaagcctgatcaacactcaccgcccagacacctttcaaacaagg  
agctaagtcaatgaggtagaaccccaaactcctccacctaggcgctgacaggcttaaagac  
15      cccattgccccacacagccctccctcctttgtaaggctcactgaggggtacaggacctgggc  
agagacccagagcaaacagaaatgaaagaacaggctttgtaccctgaagagaggaaacagg  
aggttttcaactcaaggtaactggatggcagcatttgccggcttcgagtgctgagtgaggac  
acacgtgcagaaatgacgtgagatgacacgcttagtaaaaacgatgatacactttactcgc  
acaacctgaacctctactaaaaccagccagccacaagctgtttgctatcctttattaag  
20      aggtcccacattcttgccgggactccagccaaaccagacagggtcccctaaatatagcagga  
ggcctggaggggaaggggaatgacttaggatcccaccacaccacctggaaacagaactcc  
accacagacagacggacagacggacggacaagagccggggaggagaacccacctcactct  
tggttctctccccgttgcatccactcaaaaagaaagtcaaacactggctatgcagacccc  
agccacccacccacccatagcagcgtttgtgggactccccctgaaacgggtagcccca  
25      agacaacttctatggttcttccctgactttggtttgctcctggcaactccgcgcctct  
tccttccctcagcctccagctctctctcagcatcttctaccacctactcggaccttccct  
ctctcttgctctctgctttctgggtctccctgccacgggcttcttggggaagcagcgggca  
cctttctcctagcaagggcccccactaggccctgtctgccacgctgggactcacacagcc  
gccccactctctttagggtcaggggctgagcgtgccttcgcattcgtggaggggtagtg  
30      tatggtgggtagcggggccctggcgcgtgggctgggtaagggtggggctgttggggataa  
gagttgtgcttctggggccgtaagtgtggggttctggctgtgtagaacccctccccga  
gatcggctccctccatctagggaattcctgcgaggacggtggggccttcggggacttgga  
ggctctgcggttagggggaggagggggtgcagtgacctctcagggggttggggacctggg  
gcctcaccactccccattcctggccaggactccctgtgctgaggggtgctcttgaagggg  
35      aagcgcagcttgcaatcatgcgggggtacaaagcgagctgggggcctgcgtagtcccca  
ccccggcccccgagccctgcagtccttgacgcgcagctgctcctgcttgagccagcga  
aggcgaccgcgacggaaggcagggtcctcttccatcagctctgacacacgctcccagctg  
gactgggtggtaggagggccggacagctggtagtggtcattggacactgcctcctct  
actgcctctgaccttcgggtggggcccagggtgaccaaaccagattcttcattatcatcc  
40      tcaagatcctgggtcaacgggataacccctctccatgcggagcatccggtcccgaggggcc  
tgcagctcccggtccttgctgctgttttgagcttcacttccctgcagaattcctgtcagc  
ttgtcaatatgagcccggaggtcttctacctctgccccacgggctccttccctcaccacca  
cctccaccacctccacatccttcttcttcccccacagtatcccagacatccctggccaca  
gctctccaggcgtctccaggacctcaggcttgccgtaggctccggcacagctccctcatc  
45      ttgagagcagccagagcttcaatctctgcccgtccgtgacggaagtcggccaggggcacc  
tcatagcagatctccttactgcttgcatcttcaggtcagccatgggtggccaccggggg  
tctttgcccgtggagccgcgctcgtgagggatctggtaaaactcttcgagggggccctgcgc  
ttccactgctgggcaggccacagcgttgacgatgggtctggactgtgttagggggcagc  
tcgtcccgcaaggaggaaatcagccgccagcttcttccacaagagcgttgtcagagtct  
50      tccccactgtcagaatctgcatacagccgtgctgctccagcaaaaggctcagcctcttcc  
ttttctttccggtactgattctccaagtcttgtagcctcttctccatctctagcttgatg  
tctatgccttgctgctccagccagtccttctgagcaaagttccagtcacacaggctcagag  
ggaggtcctgggggtggggggacccctcgtctcgttccagccgtgcttgctccgggtga  
ttgaagcggaacacatggttcttgcccattacaatcctgttgctgacttcagcaccagc  
55      ggctccgtcacaagcttcccattgacatatgtctcagctccttcacaaggttccaatgtg  
accatcacttctccatcaggtgagggatgctgcggaagaggcagtgctgctcccgatg  
aactggccagtcacattgatgtccacatctacctggccaaccctgggtgacgccatctttg  
atgtggtagagaaggcattcagacatcaaggggtcctcattcagatttaccagggtgggga  
gtcttttttggagagaagacacccacagtacgccatcctcccgagagcccatctcagcc  
agcaatgcttctctctccatcctcagagcttctgtcttacggagcttctcctcccaagtc  
60      tcattcagctcagctataattttctctgtttcctgcagcctctccatggcctcctcagge  
ccaatctggggctcagcactgggtgaaaatgacggctccagctcgccgttatgtggagga



5 ggagatgaggggtgaagctggggcagggggagatgatgcagcaggcagaacacctccagga  
 ctccccctctccaccttttagacctcctagagcagaggctgaaagccccctgagccatcagc  
 agttcccgcaaccggggccacctcctcctgcagctcccggataagccgggcattgggggtcc  
 tcattgatgacagcattgcatcggatctgtttgggtcgggtctgcgtacctgagagtgtg  
 agtgtctcctcgtaattgatgtcagcgggactcagggctgcaatcattgctgtgcgtgag  
 10 ttcccacccaaattctccttgagtagccaggtaagcacagagtctctgtaagggatgaag  
 tccgacttccgcttctttgattgcaaactgccagggtgagatcaccttccctagagta  
 gtcagggaacttattgatgtttgcaccttccctcagacgcatgccccgagccccctgaggag  
 tcggccccgctcgtccccggcaaggttcaccaagctgatcttactgaccttttctgaatcc  
 agtccagtaagctgggtcatgggagcgtgggtaaagacgatagtaaagacagcgtgggag  
 cggctgctgggtttcgttcattgttgggtggcagccacagttcttgccttatttccacagtcc  
 atgaggtcagcaatgtctgcataggaagtcacagccaacttagacaggtcttgtacatat  
 15 gggcccaggatgggggtgctcccggaccgcagagagccccgactcttgggggttcaagagg  
 tctcgtactcgttcgcaatagatctccatatagctcacctccacagagtaggaaagtga  
 gcactctgggttcacattaactcgagagaagaggtcctcgagagctgaggtacaatgccc  
 tgctgccccgggttccctgcccgcacctcatggtgtaggacttgccagccccctgacccg  
 taagcaaaagtgcacacgttgtagccttcaaaggcatgcagcagcatctcctctcctatg  
 tctcgatacacctgctgttgagatgcaaactgtgggtcctccaccgaagtatgtgaccag  
 20 taagaatagtgcgaatgaagcttttaaaaacatcctgctctgtttgggattaatgatggag  
 gtgggtgttgccttgcattgctgaccacacacttggcatcctggctgggtctcacgggcatta  
 aagggccgaaccctcactgccactttcacggaggcaccagccatagcttcagaatctcct  
 gccctcctcagctgggtgtcctggccccagatcagcggggctgtatcagttctggctgcc  
 ccggccctcgtatgggaagccccatcctacacttggggcctggccacaccagcaaggctc  
 25 ctcgggcagactcccggcagagagcaaaaggacaatactttgctggcgagtgtgctat  
 gaactctcgctaccgggtgaagagacgcatcggggccagttcggggctgccccgcccc  
 tcg

>scr\_gb-d79221\_3 (TOXMARKER Assignment: 54; SEQ ID NO: 54)  
 30 atgggaaaaaaagataaaccagggtgtgagcattctcgtgccgaattcgggcacgagcagc  
 attcgggaaaggcaaacagtggtctctgaagcggatgttgaatttcaatgtgcctcatgtt  
 aaaaacagtcctggagaaccgctatggaaggtactcatctatgacagatttggccaagat  
 atcatctctcctctgctgtctgtgaaggagctgagagacatgggcatcaccttgcattctc  
 cttttgactcagaccgagatccaattcgagatgttccctgcggtgtactttgtgatgcc  
 35 accgaagaaaaattgacagactgtgcccaggatcttcgaaatcagctctatgaatcctat  
 tatttaaattttattctgcgatttcaagaagtaaaactggaagacattgcaaatgcagca  
 ttggccgctaattgcagtcacacaggttgcgaaggtttttgaccagtatctcaattttatt  
 actttggaagaggacatgtttgtattatgtaataaaaataaggaacttgtttcatatcgg  
 gccattaataggccagatatcacagacacagagatggagactgttatggacactattgtt  
 40 gacagcctctctgcttttttgttacattaggtgctgttcccatcatccgatgctcaaga  
 ggaacggcagcagaaatgggtggcagtgaaactagataaaaaactgcgggagaatctaaga  
 gatgcaagaaacagcctttttactgggtgatccacttgggactggccagttcagcttccaa  
 aggcccttattagtcttgggtgacagaaacattgacttggcaacgcctctgcaccatacg  
 tggacataccaagcgtgggtacacgatgtcctggatttccacttaaacagagtaaatttg  
 45 gaagaatctacaggagtggaaaaatttccaactgggtgctagaccaaagaggaaaaacaag  
 aagtcttacgatttaactccagttgataaattttggcagaaacataaagggaagtccattc  
 ccagaagtgcgagaatcagtcacaacaagaactagaatcttacagagcacagaagatgag  
 gtcaaacgactgaagagcattatgggcctagaaggagaggacgaaggagccatcagcatg  
 ctttctgataaactgctaagctcacatcagctgtcagttctttgccagaactccttgaa  
 50 aaaaaaagacttatcgatctccatacaaatgtcgccactgctgttttagaacacataaag  
 gcaagaaaaactggatgtatatattttgaatatgaagaaaaataatgagcaagactactctg  
 gataagtcctctctcgacgtcatatctgacctgacgcagggaactccggaagacaaaatg  
 aggtgtttcttatctactacataagcgtcagcaggcaccatctgaggttgatttggag  
 cagtataaaaaggcttttaacagatgcaggatgcaaccttagccctttacagtatatcaaa  
 55 cagtggaaaggcttttgccaagatggcctcaactcctgccagctacggaaacactaccact  
 aaaccaatgggtctcttgtcccagatcatgaatacaggatcccagtttgtgatggaaggc  
 gtcaagaacctggattgaagcagcagaatctacctgttactcggatttttagacaatctc  
 atggagatgaagtcaaaccgagactgatgattacagatatatttgatcccaaatgctg  
 60 cggagcaatgcagcgtcagttcctaggaacaaaagtcattccaagaggccattgtcttt  
 gtggtaggaggaggcaactatattgagatcagaatcttgttgactacataaagggaag  
 caaggcaagcatattttgtatggctgcagtgagatttttaatgctacacagttcataaaa  
 cagctgtcacagcttggacaaaagtaacacagaagagtcataatgggtgatcagtggtgga

cagatgtaaaaagccagacgtgtccttctccatagcagtgcctaacagtgcaacctgcg  
gaatcagtcatttttaagaaattctatacttcatatactgtacaatgattaaaataata  
aaccatttcagaagtaaaaaaaaaaaaaaaaaaccc

5 >m61937 (TOXMARKER Assignment: 55; SEQ ID NO: 55)  
ctcaggtttctcacactcctggtaatactgtaaaactttaccatggaccacagttccaag  
gactcctgaacacagtccttgagtttaagcctgtgaacagcccacgcttcccatcgatgcg  
taacaagcgatggattccatattctctgcgtgtagcactaaatgatggtaacttcattcct  
gtactgggggtttggaaccactgtgcctgagaaggttgctaaggatgaagttatcaaggct  
10 actaaaaatagctatagataatggatttcgcctatttgactctgcttatttgtacgaagta  
gaagaggaagtgggccaagccatttagaagcaagattgaagacggcactgtgaagagagaa  
gatatttctatacttcaaagctttggagcactttccatagaccagagctgggtccgaact  
tgcttgaaaaagacactgaaaagcactcaactggactatgtggatctttatattatcat  
ttcccaatggctttgcagcctggagatataattttccacagagatgagcatggaaaacta  
15 ttggttgaaacagtggtatctgtgacacatgggaggccatggaaaagtgtgaaggatgca  
ggattggccaagtctattgggggtgtccaactttaactgcaggcagctggagaggattctg  
aataagccagggctcaaatacaagcctgtgtgcaaccaggtggaatgtcacctttatctc  
aaccagagcaaaatgctggactattgtaagtcaaaagacatcattctgggttctctagtc  
acgctgggaagtttcacgagacaaaacatgggtggatcagaaaagtcagttctctctagc  
20 gatccagttctttgtgccatagcaagaagtacaagcaaaccccagccctagtgtgccctt  
cgctaccagctgcagcgtgggggtgtgcccctgatcaggagtttcaacgcgaagcggatc  
aaagagctaacacaggtttttgaattccagttggcttcagaggacatgaaagccctggat  
ggcttgaacagaaatttcagatacaacaatgcaaaatattttgatgaccatcccaatcat  
ccatttactgatgaatagtaacatgggtggactttgtcagcatttctatcggaagatctgt  
25 ttatgcattgtgatttgaaagatatcttgatactgggtgactgaatgcatcagaccactg  
tttctgttaattcacagtcagctggagcaatgtccacagtgctatgaggggaagccatgtt  
tttgtcacactctgaaatggaacatcacgttgcttttcttctgtgtttttaaatattcatt  
tattttgtcttccatataatgaatttttccctacatgtatgtgtatctcatgaatgtcta  
tgctcatgcaagggtgaagagtgtgaggtcacttggaaacggagttacattgattatg  
30 gaggtaaccatgtgggtgctgggagccaaacctagggtcttctgtgagactagcaagtgcct  
ttgaatgctgagccatctcattaggtccaacctaaagatccttgctgcccactatttct  
gtgatctcaatgttttgttttctcctgacttctgacaccaagctgatttgctagaagtct  
tgggcatgaagtgggtgttgaggacagttattgcaaagggatttctgggtgggagttgaa  
agaacgttcaacattcaggggaattaattgttcgaggttattgattagtcattatccccc

35 >cszr\_229602935\_183895355 (TOXMARKER Assignment: 56; SEQ ID NO: 56)  
gtgcaacttgtccgaggcacctttgcagacacagccctgggcacatttgagcagcccacg  
gggcagcaggagcagcagctcttctgcaggaggtgcatttgagttcttgagcagccgag  
40 gagctggagcaggtgcaggagccgcccgtggagcaggagcagttgggggtccattccgaga  
tctggtgaatctggagcaacggtgtaagcgacaagaaggcagttttttttttttttttt  
taaaataaacaggtttttattttccacctgctcggtacaaaacgggggtttattaaactgg  
gtggaggtgtacggcaagactctgagttggtccgga

45 >scr\_gb-af106944\_3 (TOXMARKER Assignment: 57; SEQ ID NO: 57)  
ttttttttttttttttttttccaaaacaaattcttttataagttgtcttgtcatgttttgtc  
acagcagaaaagaaaagccactaagacacttgctaatacccccggttctgtttttttttctca  
aaacccaagatatatatatatatatatttacactcattttacatatgcaaaaaata  
gaaccagactcttctccctaagacttccctgaaaaacctaactcagaaccttgcaagtac  
ctgatttctgtttatttgagcttctcttccagaatcaagggaataaagacaaaaggtttatt  
50 tttcttactccaatgctccaggaccaacctggcatgggttttcattccaggagctagca  
aaataagggtatgaaagtttaggtatcttgctgctgaatttcagtttccctaagggtggaga  
cagctccgtgtaaatgccagtaaacaggtacttggtgagctaaagtcataaaggaggag  
cagtgcgccagaataaattgacagttaatgatgtcaagtatcttaatgtttatttttatt  
ctttacatccagcacttgaagaaaaagaaatgacatagtggttttagaaacatagtccttc  
55 atgattataactcatcaataccttagaacacacaaggacactgtgagttaatgactacac  
taaaaaataatgggaatttcagcataatttaacaaaaatccaagaggaaatttcaggacct  
tgatcagaagctttcactaagtgctggcactatagctacttcatcttactaagtgctgg  
cgctatgtgctacttcatcttactaagtgctgggtgctatgtgctacttactgtagacca  
agcttcaggggcaggctaagaaatcttaacctctgaagacatgatctaagaaatggggac  
60 caagcacttgtagagaatttggtagccatcaagaagtccttagtaaggacagctatggaag

gagctggccacctttaacctgaacctgtctttaaattacaaagcccatggagcagtactt  
 ataaacacaaagcatggtgaggttttgccattctataaataatcttcaggattccagctgg  
 ggctctcttttggcatgagaagcttcaggtaaaccagcagacataggatgacctattatt  
 gatggaccttctcaaagtactcttttgaagctgttggacttggcttgatcgtaggggact  
 ctggtgtccagttgggtgggcagacttctccatgggtctccacaaactggaacgccttta  
 ccaaacggagtggttcttccacacttcggcccaccggaaggtcattgacactcaggtgct  
 tgatgacaccattaggggtcaataatgaagagacctctgagcgcaatgccagcactttcca  
 acagtactccgtagtctcgggatatactgcttagttaagtccgacaacagcgtgatgttca  
 tgtggcccaaaccaccattctttcttggcgtgttgatccaggcaagatgactgaagtggg  
 aatccacagaaaccgcaactacttccacagtttacgtcatgaaactcattggctttgtcac  
 tgaagcaacaatttctgtaggacacacaaatgtgaaatccaaagggtagaagaaagca  
 ccaagtatttcccccttaaagtcgtcgagactcagctctttgaaactctccattgacaacag  
 cagtacctttaaaatggggcgcatgctgggtgacagcaggggtgtggaatgaagaactgg  
 tgctaaaggcaaaccttcttggggacaggcagaccacagcatgtctgtcaagcaggttc  
 ttctagaagcaacaggcctaagaactgttgaggcagaaatactccggaaaatagtgtctcg  
 caggccgagccaccgaggaccagagcaaccttcccgagctgccgccatcttcagagaac  
 gcaagagccacgatagc

>scr\_gb-m11794\_3 (TOXMARKER Assignment: 58; SEQ ID NO: 58)  
 atgccacatttgtgaccagtacatgtttctgccaccatgttcgagactatcaaagtcc  
 agaggggtcatcaatccacttatcccaaataaggtgcaccaatcccatttcaacgcctc  
 tgccagcccccttatttccaatgaacacagacaaagctgggttaatacagtcaagttttt  
 tattttattgtcagttacatgctttatagaaaaaagtgtggagaaccgggtcagggttgta  
 caaaaaaaggctaggttccctacgttgttttatttacaccattgtgaggacgccccact  
 tcaggcgagcagctgcacttgtccgaagcctctttgcagatgcagccctgggagcactt  
 cgcacagcccacggggcagcaggaacagcagcttttctgcaggaggtgcatttgcattg  
 tttgcatttgcaggagccagcgcaggagcaggatccatctgtggcacaggagcagttggg  
 gtccatggcgcaatggaggcggttgagatcaacgagagatcgctgtagagttctagg  
 agcgtgatggagagaagcacgcggagcgcgacctttatagcccagagtattgggtcgcg  
 gcaaaagctccgcccgggtggcggggcgccacctgcccctctcccactgctgcacacg  
 ccttcttctggtcaagggaatgg

>scr\_gb-af069306\_1 (TOXMARKER Assignment: 59; SEQ ID NO: 59)  
 ttttttttttttttttaggaaaagcgactgctttaatgaattagacaaaatttcacat  
 gaaatcagaatcctataatccttcccttctgatcactaaaaaatgcaagattcattcggt  
 acaagccatgtgcgattcggacccctcgaaggcagtgaggtctgcggtccagcctcagg  
 tctgcactatttccattctcagcgctgaacattcgttctgtgagcatccgctccaact  
 ttatggcatcagcagcaaaccttgcggatcccatcagagagcttctccacagccatttgg  
 cctcattgtgcagccaacggaaggccttctcgtccagatgtatcttctccaagtactgg  
 tctgggctgctttgacggaagcggtgggtgccagcttgcctgtctccttgagcagctccc  
 ccagaagcttgggtgagatgggtgaggaaatcacagcctgccagcgcttgcctcaccgg  
 tgttacggaagggaagcaccatgacaatggtctttagtgcacaaactttttagtagttagt  
 agatttttgtgacactcttccccagggtcctcctggggttcgtaggatttcttgtctg  
 tgtttgccacatgccagtcaggatgcgccccacaaaggagagatgagcgtcacgcccg  
 cttcagcgaggccacggcctggggaaggagaaagcagtgatgttgcagtggtgagc  
 catgctgctcctccagctccttccggcctggattccctcccagggtgatgataacttga  
 tgagaattctgtccttgcctgatccagcttcttgtaaagctctatgatgcgcctggctc  
 gggccaccatggcatccttatcaaaggaaagccttgcatcgacttctgtggatacacggc  
 ctagaatcttctttagtatttctgccccaaacagcacaaaaagtttatcaatggcatttt  
 taatctgctcctcttgtggcccaaccagcttcttgccgtaggcaatggcctcctccacca  
 gctcttggtaggcaggcatctgtgctgcagccaggatcagggtgggttgggtgggtgagc  
 cctggggcttgtactcatcgatggcggttgaaatcaccggtgcagccaccaggtgggtga  
 actgcttgagctgggtccaaggcggaactccatcctctggcgctttaccggggaccgccgaca  
 tggcgaaacgcgcacagctgaggcggttagctggt

>scr\_gb-d17310\_4 (TOXMARKER Assignment: 60; SEQ ID NO: 60)  
 gcactctccagcctctcaccgacttttttttcaaggagacaattttatttttttaccac  
 gctgaattttataaccataacatgggttaacagagggaagggggaagtgtgaaacattttacac  
 aggccaaagggcacagtatacatgtagtcagctgatgtcaacaggatgttgggtttttcaga  
 aagcttacagggtcatcacattgggtatcttgatgtcagatgtatttctcagcaagggtcag

5 aactttatcatatcattattcactcctgaccaccagatttgtattagtcttctgcagctgg  
 ctggggattttccatgaaccagctcacttaatttctaaccataacatcaataatggagg  
 gtttcaagggcattgtctccaacatgtaattacaaaaagaaaaagatgatattttccc  
 10 aaaaagagagacacattcaaatttctctcaaactccccacatctgaatcatgatgatgc  
 ttttaaattgggttcttcttaccacattccaaccttcccacaagaacttgctctccag  
 gttcttggagctctggttcttgggctggtggagagaaccctgggtctcttgggtcactcct  
 gccacaggtgccctacctcaaaactaagaaaaagggaatctatggagtactttcttct  
 tctcaagaatattgggaatattgactaatcaataacctcgaacaattaattccctgaa  
 15 tgttgaacgttcttcaactcccacccagaaatcccttgcataaactgtcctcaacacc  
 cacttcatgcccaagacttctagcaaatcagcttgggtgtcagaagtcaggagaaaaacaaa  
 acattgagatcacagaaatagtgaggcaggatctttagggttggacctaatgagatg  
 gctcagcattcaaaggcacttgctagtctcacagaagacctaggttgggtcccagcacc  
 cacatggttaactccataatcaatgtaactccggttccaagtgaacctgcaacactcttcaa  
 cctgcatggacatagacattcatgagatacacatacatgtagggaatattcatatat  
 20 ggaaagcaaaataaatgaatatttaaaaacacaaggaaaagcaacgtgatgttccatttc  
 agagtgtgacaaaaacatggcttccctcatagcactgtggacattgctccagctgactgt  
 gaattaacagaaacagtgggtctgactgactcagtcaccagtatccaagatatctttcaaa  
 tcacaatgcataaacagatcttccgatagaaatgtgacaaagtcaccatgttactatt  
 catcagtaaatggatgattgggatgggtcatcaaaatatttgcattgttgtatctgaaat  
 25 ttctgttcaagccatccagggttctcatgtcctctgaagccaaactggaattcaaaacct  
 gtgttagctctttgatccgttctcggttgaactcctgatcaggggcacaccccacgct  
 gcagctggtagcgaagggaactagggctgggttgggttgaacttcttgcattggtgac  
 aaagaactggatcatctaggagaactggacttttctgatccacccatgttttgcctcgtg  
 aacttcccagcgtgcagtaggaaaccagaatgatgtcttttgacttacaatagtccagca  
 30 ttttgcctcgttggatgataaagggtgacattccacctgggtgcacacaggcttgtatttga  
 gccctggcttattcagaatcttccagctgcctgcggttaaagttggacaccccgatgg  
 acttggccaatcctgcactcttccatggtcctccatgtgtcacagatatcca  
 ctgtttcaaacaatagttttccatgctcatctcgtgggaaaaatataatctccaggctgca  
 aagccattgggaaatgaataatataaagatccacatagtcagttgagtgttttcagtgt  
 35 tcttttccaagcaagtccggaccagctctggtctatggaaagtgtccaaagcttgaag  
 tatagaatatatcttctcttccacagtgccgtcttcaatcttgccttcaatggcttggc  
 ccacttctcttctacttctgacaaataagcagagtcaaaatggcggaatccattatcta  
 tagtcttttagtagccttgataaacttcatccttagcaaccttctcaggcacagtggttc  
 caaaccccgatcacaggaatgaagttaccatcatttagtgctacacgcagagatatggaat  
 40 ccatcgttgttactcatgcaaccaagcaggtcttgggtctggcgagggtcttctgactg  
 ttctgagacagccctgtgtgaggaatgcactttcacagggttggaggtacttccaagacg  
 ccataggaaccacagctgggtcacagctatcagttcactgtgggcaagaaacctctttat  
 ggccacctggtaacaaaaatttttctgtctgtgaatttttcttactattttaa  
 45 >scr\_gb-bf281368\_2 (TOXMARKER Assignment: 61; SEQ ID NO: 61)  
 tttttttttttttttttttttcacacaggggttgcctttattttccacatccaacttgagcagag  
 gccctgccacaactgaacagctgtgaggtgctgggtgcctccagagtttctggcacagt  
 aagtggtcagggtgtgcagacttctgatggccacatgacactggccacacaggaacagca  
 50 agtccatgaatggaatcccactgagctggaagtggaggctctggaaacccatgggcag  
 cagcaggagttaaaggagccaccaggaacactgcagtgagggtccaatgcagacagggct  
 gataaaaacccaaacagggcattgtgagagcagaggctcgagtgtcccgcgtgaggacc  
 ggggctgaaggcacagagctgtgtcgggatggaagaacctgggtgcactcgagctccag  
 agcacgaaagcacaggtgagaacccagcccagggtctctgtgaagagtgtggccttggg  
 55 tcttgggcacggcacagtgacacacagtgctgaggtcactcctgacttcccagaggaatg  
 acctcttcagtgcacaaaaactcaatgggtctcttctccagtcacacagttgctgtc  
 agctcgtcagtggtccacccctcccgtagctctagacgctcgttcttctgcttcatatag  
 agttcctgggccatttttccggtattgctgaagtctccatcatgggtccgccttcttcc  
 accagttccttgaagcttggactgggtcaagcgatccttctgctcaaagatcttagag  
 60 tatttcttcagatccttttaatttgctttatctgatcctgactgaggagtgttgggggc  
 cttgggtctccagagcagctggcagaagcggtccttgtgttcttctggagaagacgacct  
 tgggaaggtccacagccaataagcattgtccaccttatgggtccaccacgacacagaggt  
 accacatagcggcagttgggtcccatcagcgtcggaggccatgtagtgtctgcaatg  
 tcatcaggtgcagctctgaagtgtcgacaaacgccaaggcgccattcatgctcctcagc  
 ctgctg

60 >scr\_gb-u56407\_3 (TOXMARKER Assignment: 62; SEQ ID NO: 62)

cccaacacaaacagccagccagccgctcggtgctcgccgctccacacagggcccgcc  
aagacaaccttccaccatggcctttgaagagaatccacaaggaactgaacgacctggcgca  
ggatccccccagcacagtgttcagcaggtcctgtcggggaagatatgttccattggcaagc  
5 tacaatcatggggccaaatgacagtcctaccaggggtggagcattttcttgacaattga  
tttcccaacagagtaccccttcaaaccacctaagggttgaatttacaacaagaatttatca  
tccaaatgttaacagtaaatggcagcatttgtccttgatattcttcggtcacagtggtctcc  
agcataactattttcaaaagtacttttgtccatcagttctctgttgtgtgatcccaatcc  
cgatgatcccttagtgctgagattgtcagatctacaaaacagatagagacaagtacaa  
10 cagaacagctcggaatggactcagaagtatgccatgtgactaaagagattattggatcc  
tctgcaataaaaagctaggggaactctgaaagagaaagtcccttttgattcccacttgact  
gtttgctgtgaacccacgatgtaccggcctcgctcctcctgggtgcacggcttcatctg  
tacagtactgttgcatgttgacgcacccaaaaatactgtgttctgtaccaacactgtct  
cctagcagacgagccttctccaggcataacctagggtgtgagattaaaagttttcttatt  
gacttaactctggaataaacaggtgtgagtgagggtgggtgggtacagataactgtcagaa  
15 ggggtaaaagttcccaacctataagacaatgagatgggtggttttctgtggaagccatttaca  
gctaaatgtttaaatgaatgaaaagctagggtgaagaacatgaatgttctgtactcattt  
tattccaaaagacctagagcttaaatgaacattaaagccaaccagactaagccaacccac  
ctcctgtattttaaagtctaattgggtcaacaaaaatagatcggcactatcgggtccataaa  
20 gtgtgcctggctttgttcccaaatcctttatacacggatgactcaacctattttctttca  
cactttctctccatattcttttggtttacttgcggtttctcagttgattcatcactaatag  
ctcttatttttattatttaactgtctaatctatttggatgtaaaaggtagacattcaact  
tgtgaaaaaagcttgtgtatagagacctaatgtctcctctggagcttgtacagtcagg  
aatgatgcatctgtgtaataaaccaattattctagccattat

25 >scr\_gb-ai406674\_1 (TOXMARKER Assignment: 63; SEQ ID NO: 63)  
tgtacactaccctcacaaaccacaagccgcagcaacatggatgccagctctggagcagc  
aacagccaggatgacctggagccaggggggcttcggaacagatgtgcaccttctctgggt  
gatgttttcagctttgtgagaaaccttactatcagaggagatggctagcaatgttaccaa  
30 caagacagatcctcgatccatgaattcccggtgtattcttggaatctcaacactctggt  
ggctcaagaagctctgatgtggaggccatctttcaaagtatggccaaaatttgtgggtgtct  
tgtgcataaagggtttgcctttgtccagtatgttaatgaaagaaatgcccgagctgctgt  
agctggagaggatggcagaatgattgctggccagggttttagatattaaacctggctgcaga  
gccaaaagtgaaccgaggaagagcgggtgtgaaacgatctgcagcggagatgtacgggtc  
35 ctcatcttgacttggactatgactttcaacgcgattattatgacaggatgtacagttacc  
agcacgtgttctcctcctcctcctccattgctcgagctgtgggtgccttccaaacgcgacg  
tgtgtcggggaacacctcacgaaggggcaaaagtggatttcaattcaaaagatggacaac  
gggtatcttcttccaaatctggaaagtgaaggtgatgcacttcaggcttaaaaaagga  
gctgactcagataaaaacaaaagtggattctctgctggaaagcctggaaaaaataaaaaa  
40 aaaaacccctcgtgcc

>scr\_gb-bf290678\_2 (TOXMARKER Assignment: 64; SEQ ID NO: 64)  
ttttttttttttttttttgattttggccaaactttttatttagtattttgtagtgttta  
acacacacttaaatggcttactcggggagggggaagggaggttcttgtanattcccaag  
45 gaaaggtcagaaaaagcaaatatggccagcatccatttctgttttttgaggggggggggt  
ttctgggttaaatagtacatcgcttaggcattctgatctcagcttgggtttgtttgttgaata  
tatataactgcgaacattgagatttcagttggaagacaccttgaaatcctcacaccca  
ccaacctctctaattggctagcttgtctgcacaggcaggggtgattcaactctcaatggag  
accaaaggacatctagatggctaaatgtttgtggaagatcttgggggtgtcttgccctcatt  
50 tgctgggaaaaatcaggaagtggccttcagggacacttttacttggaaaattacaacact  
agttacaagtacagggttacacatctaacatttgtttgttgaagcaactcataatagca  
aataaaattaaacatgtcttacttttccctcacagaacataaaaaattattaaggggaa  
caggaaatttttaaaaaggtaacacaatttttctttagtagtccttgggtagtttatgac  
agaaagtttccatttttttgtttgtttctttgaatggggattgttgggtccctcgtg

55 >scr\_gb-bi288503\_1 (TOXMARKER Assignment: 65; SEQ ID NO: 65)  
tgtacagttgctagtttgaggctgggtgttgatgttctgacaagagtggctcagccatggc  
tcagtagagtctcttctggaagtttgagaaattctggcttacgggaaaagggtttttctt  
tcttttcaagatatgtccaacaaagtctcttcggtcagtaatttctgcagtcagcctt  
tcgtccgtctgtcagcaaacctccaatgcacacttgggagtcagtcataaagggttaa  
60 gcgcacacaagcgtggccaaactagtaggtccgagaggttcaccggcaggccaccgtactta

atatgcagaggggtgggcttcacgcctccccgcgagcgctccacgggtcgaggagtgg  
tgggcaaggagatgagggttaagtccaatgggttaaaaccaaccccgagagggttaaaac  
taccgatgacgctgccacggagggggccgaatccac

5 >scr\_gb-d86383\_2 (TOXMARKER Assignment: 66; SEQ ID NO: 66)  
ttttttttttttttttttttcacaccagatgacgaatgtatatgaaagttaattcattaaat  
taaaaaaaaaaatcaaacatttggggagggttttttttacaacgaataattctatacaca  
tgctatagacacgggtttctataaaacacactatctacaatctacttacatttaattgtcc  
10 tgctatttctagttcatgtgagatcagtcacaagtgaagtcagtttccctgcctgtagaga  
ctgcgtcatcccttaataaccagggtcagaggcactggccgagcaaaacaagattgtaaga  
atcttatcaactatcttgccttatgagaacagacaccagggggccaagtgtctgaaccggc  
tttgaggttaaggcagcaatgtaagggtgtcacgtaaaaaccaagtgtgtctttgaaagc  
attccatggatccccaaatgctggccccctttctaagtgcacctctgaagtcgagggaac  
15 agctacacatttgggaaaagtcattcgagaacagccgccccaaacctttaaagttatagt  
ttaagcttcaggcaaaagttcaaattacttctcacaatagaaagaattcactttttaaa  
aacgaagtcacatttagccactttatcaaaacaacttaacacccggtacggaaaacgtacg  
ctaaaccaaagataggtttcaatgcacgcccgtgccaaatattttcaaacgctagaaga  
atggtacttcttctctcagaatttcccagtttgcctgtagcagaacggtattctaaagt  
20 ccagtcctcgaacatggtcacggccgatgactgtcatccagcattaaaatagcctttatc  
accctcgatgtccacttccctgggtcggaatcctctgagatctctgattcagggtcttccgg  
agaggctggggagggtgaacactgagaactgtccaaagaggcacctttattctgttccact  
gggcaagtcttggccctgggtcacaggaagtgtccaaactgtccaactcatcctttttatt  
gctttgaggattctcctgcttcagtcgtctccatttagctctgcgattctgaaaccagggt  
25 tttgacctgtctctcgtgagctgcaacatcttggccagacgctttctctcagggtgggga  
gaggatttctgagctcgaacttcttctccagctcgatggtctggtcgttggaaaacct  
cacttgaccgcttcttcttctgtgcagaggtcgtctgtaggaaggggtccagagcaagggt  
cttgcccagggggtcgtggcggagtagggcgtgctgtagtcggtcaccgctccgaggga  
30 cgggtacagaggggtccgaagccaccggggccataggcagcggccagcgcggcgggg  
gtgatgcgagaaggcggggtggaccggcgtgggctcgtacaccggggtccggtaggagga  
cacgaggctggtgaaggaggagttgggggacggcagcgtgggagtgggcgtgggagcggc  
gggcccgcgaccaggtgctcgtcgatgtagaaaggcgtcgggtgagcgggctgcagcag  
cggcgtgggcgctacagcgggactccgacggcgggcgagccgcgggccccgggtgcgg  
gaactgcattggtcgtccgc

35 >scr\_sc-133366194\_1 (TOXMARKER Assignment: 67; SEQ ID NO: 67)  
gctagcatcttttttctgccacgaggtgcgtttttattttcatcaatcatacaaatgattt  
tccatatcacagggaagctgagtgccctgggtgtgttcacagtgtagcttgtcgttgtg  
tctgtccatcttccccgtcagaatgggtctcagaaatgatgaggtgaggtggagaaatc  
40 ctctaggctttaggaaattttactcctcttttctgttgaaatggtcttttgggtgggt  
gggtgttcttctcatgctcttgggttttctccagtggtgtttattgaagcttgtgatttc  
ccccatggataacttgctgccattttcttagaactcttggaatcttgcctctgagctcat  
gtcccaattg

45 >cszr\_230290139\_182026368 (TOXMARKER Assignment: 68; SEQ ID NO: 68)  
aagctttggagctgctaggtgctacctatgtcgataagaaaaaggatctgcttggagccc  
tgaagcattggagacgggcaatggaactccgccaccagggtggggactaccttctaagc  
ctgaaccccagcaactggttctagcctatgactattccaggaggtgagcacgccccaa  
50 agttggaagccctcatcacagatcctgatgagatgcggatgcaggcactgctgatacggg  
agaggatcc

55 >scr\_gb-ai013477\_2 (TOXMARKER Assignment: 69; SEQ ID NO: 69)  
tttttttttttttttaagaagctgttctcatctatgaaccagatggcatctaccccat  
ctgttggtgatcagtcgatctttatgccactcctgtgctttagtgagcacctgggtgac  
agtcagatgaggggtgtctaggtcagggtccgggagcagggtttaggggttagactcg  
60 taggggcagctctgggatcacaaggaacaagtgggatacccgccccacgccaaggtccacc  
gttcttccgggtagtcctatgagtatcatttgttgcagtagcccttctactcaaggtctt  
ttgcttgacactagccatttggacgtaggagcacagagtgttgggcccccgattcaca  
caacaactgggcgggcttcccttctatctttttgcatagccagcactctaggaccaagag  
gcttgcttccagggtgctggagagggccctcttgttcttccctggggcagtccttgacc  
agtgctcctttttcttgcattaggcacactgatctttagccaggaattctcttctgttgc



aggccccatgtttctcctgtggaaagaggggtgtggggctctgggaaaaggccactcttca  
aacattcatga

5 >cszr\_204229614\_191891958 (TOXMARKER Assignment: 73; SEQ ID NO: 73)  
gctagccttatgccagcctgccactgtcaacatattctgttccattgggtacatgcttg  
atacatacactctgtgtttttggctaattgagctttttaattctattgtaatattttca  
attg

10 >cszr\_204229615\_191892510 (TOXMARKER Assignment: 74; SEQ ID NO: 74)  
caattgaaaatattacaatagaattaaaaatctcaattagccaaaaacacaagagtgtat  
gtatcaagcatgtaaccaatgggaacagaatatgttgacagtggcaggctggcataaggc  
tagc

15 >scr\_gb-aa801331\_1 (TOXMARKER Assignment: 75; SEQ ID NO: 75)  
ttttttttttttttatgaagacacgaaatgcatttattcacataacaaaaacaaaaac  
aaaaacgaaaaaacactcactccctcttctacttgaaatgtgtcagtaatgactcaaagt  
gtcatgatttaccaggtgggtgaattcttctgacaaccaggtgaagaattaggaaaacata  
cagttccagctctttatattctgaccctagaaatcggttcatttgtagctcttgggggtac  
20 acagtaaaagcaggcaagcaactgtccacactgtttcattccacataacttagtgagtgccc  
ttattcagggcctaacttctactccaggcacaaaaacaaggcaggattgcctggtaagtct  
gaacatgagaaaagaaaaagatttattacacaacagatatatccatttatgtgagtggtg  
acatctaggaattctctgctttatagacaattagaagcagcatcctttctttagaatatt  
tctatgccctcactaaacccatgagtaagtatcttgcctgggagtcatacccgagctaa  
25 ttacaattcaatattctccctgtacatgcaatccttgaaaaacggttatatgtattttatc  
tcattttcataaaaagaattacaaagacccccaaaaagggttagtggttggttgcatattaa  
ggttgcaattctccagaaacccaaagttcggatagtagtgacggttggtgcagacaatagt  
ttacctcatgctacaggctataaatgtcagaacagagcttaaacactcacattagtgaac  
gcattggcactacttgtactctttatttttaagggctaagaaaaagcacacttctactcag  
30 cctatggaagttatcagtgagcacattctctatcgctcactgtacagtaaaactatgtac  
aacaggcactataacaaacagaatttttagagtcaggtatgacatgaaactttttcaattt  
tttatattttacactgtgggtttatcctcatcttaagatcagtttttcattttgtttgtt  
cttctgtttttttgtttttttctgcctaaacgggtatgctcaagtagcatggataaatc  
ttccagaatatgactgagtaactccttggctcttcccagagccttgccctcagcacagc  
35 atgatgtttaaagatgggtctcattgtagacatcaaagtaggtagaagaacaattgtgtct  
gtatcagaggctctatgaagagacctggagtcctgaagttccttcttactac

>scr\_gb-aa899865\_3 (TOXMARKER Assignment: 76; SEQ ID NO: 76)  
aaattttggctggatccacagcaagagtcctcagtattattttgtttgtttgtgt  
40 tctgttttgtttttactgcaacttgacaataaaagatgtttggcattggaagagaataga  
acattaggtctgggccagcgcctctgactccgtctgtttaatagtttaacctgaagtcg  
caagactgggataaaacaggagagctgacatgaaggacatcatcgcacatgtttcggtta  
ctgtgtcagaactacacgtgcttggccttatttcttggagcctgtggcagaagagtgtat  
cgaggcagaaaagcagaaaaggtccaacctccctttctagaaaatgtcccttgatgtcctga  
45 ttctcttccactagtctcactactaaaggtcctgtcacctctcagtaaacactgtgggcc  
gggaaccaagcctcagggcacaggcctttgggagctgtattagagttctcgagaagaata  
aacagcacttgacagaaggaattcccagaagaagaatgacttacaggcttctgtccagcta  
atccaacagtgagcagaaaagtccaaaaatccagcagttcggggccatgaggctgggtgtct  
cggtggttcttcagtagactctggaatcccaatgacgttaggctctaacgccagtgaaagga  
50 atggacttgccaacaaggtgaggccaagcaggcagaagagcaaaagctcccttcgtcctgt  
cctcaagtagacttctagcagaaggcgtggccagactagaggtgtgtcttcccacctca  
agatcaggattaaagaagatctactgacttcaaattaagcaaaactccctcccaggtgtg  
ccctctgtcatttagatttttagttcattcaagatggagtcaagttgacaaccaagaatggc  
catcacgggggacactccacatataaactgtataccaagcttcataatttcagacatgttt  
cttaatgtcatccacgtctccagccctgtagtgtgtatgtgttattctctgcagaat  
55 ttagcatgcccggtgtttcctgtccttcacataaacgcctttgtgtgaagcttgcttgatc  
ctccactccctctccagccccacccctgtgacactgccagtaataactgttcgttgt  
ttacctgttgcttgaagtgaagtattaaagcaatttgaaagctaaactcacctgtaag  
actataataaatacctgtaatccaataaaaaaaaaaaaaaaaaa

60 >scr\_gb-aa997629\_1 (TOXMARKER Assignment: 77; SEQ ID NO: 77)  
tttttttttttttttaataataatgttactgtcgtgttggtgtatattcattgcata



tacttcaggaaaagttttctgttcttgctaaataacaaagcacaattggttaagttccat  
 ggacagcaggtcctcctcagaacgtagccagttctgtgagggcaccatatacccaaggaca  
 agcttgtggcatgccagatgaacagcagccttggttacacgcacacctgtacataaaaag  
 ctcatctttccaaccacgtgcagccaagagattaccacagacttgacacagggaccctaa  
 5 caggctcctatagacagtcctgccgctccatgaagtggggaaggaacaaatgcagtgacc  
 gcatctaatgcacttcctttgaaaatgtttgcttat

>scr\_gb-aa997691\_1 (TOXMARKER Assignment: 78; SEQ ID NO: 78)  
 agatctgagcggccgcccacggctcctgatgacagaagagctctcctccccgaaaggggca  
 10 gtccggagcccaccagtggtataccgccaggaaggagataaaggcagctgagcacaatggt  
 gctccagaacgcacagaggatgaggacaccggagcccctggaggagggtctagcagag  
 gaagctggcagggtgagcgcagtgacagcaggggcagcccacagggtggccggcgctat  
 gtgcaggtgatgggcagcgggtgctggcggagatgaaggctaaacaggagcggagagca  
 gcatgtgcgcagaagaagcttggaacgatgtcatctcccaggatccctccagcccagtc  
 15 atgagcaacacagagcagattagatggaggggcaacagtgccctaaactgcaaccaggctct  
 ccagagggcccgctttgggttggaacaccagaaaagaatgccaaagctgaacc

>scr\_gb-ai411514\_4 (TOXMARKER Assignment: 79; SEQ ID NO: 79)  
 tttttttttttttgttggtaaggtacagaaaactttattggaaatctcttgattatat  
 20 ttccaagtgtagctctcatctcctaacaagcactggaggaggggcttcacagccacctg  
 gtcccagcctgagcttggtgctggggagttgtctagagccggtttcttccattgtgtagga  
 ctgaggggcacagggccacctgaaggatgcttcgctcagcttccctggcctctttcttaa  
 gaatctgggacataaaggctgctgtctagaggccactggctgagccctgaaaagaatccg  
 tgccctcaccctcttttagtgctggccctggggggtaaatcctgttcagtaggctatga  
 25 atgtgccctgacccaaaggctgcaatggcacttgccaccactgctgggcacatttctc  
 tgtggcagcaaaagcatgcacaggggaaaggctccagtggttacatgcagattactaacag  
 cagttgagagccacctgctccaatgcgtaacggctgctgccagtgaggatccagggacaa  
 gaacaggacaggctggcagaggcacttgactgactcaagcaacaatacctgaaggtttaa  
 gtcaaccataggctcagctttgggtttctcaaaagggaaccaatccagcttgtaagcccag  
 30 ggccatgtacagactctggaattagaggaggagagaggagggaacagctccctagctcc  
 tgctccagctcaggggctggagcagcaggttatacagtgtcctctgggcaccatgggca  
 acacacctctgaggagtcctcacactgaacacacctgagacctcctgggctgctagaaca  
 gagctagtacattacagatgctgtgtcaacagagtatgctcggcaggagcacgcagcat  
 gccgggaagcctgatgcctgctcagttccatacacacagtttgagggggctactttgcct  
 35 ttggcagacccttgctgtatctctccttaggtgtgacaggaagatcctcagagcagtagc  
 acagggttctgagtaatcttcaccggaggcctacagcccagagaaaccctcctcctcccc  
 agcagaactgctaaccctaaacatacttcttttataaaatatctgatttctctgacagta  
 ataaatatttaccatgttctatatccacgcagcagcgatcgagggaaaacgaggaggaaa  
 40 aaagatcctacagcgccgcg

>scr\_gb-aw142560\_3 (TOXMARKER Assignment: 80; SEQ ID NO: 80)  
 ttttttttttttttgattgaaaatgtttaatttgtaaggcacacagtttatgatca  
 ttttaatatctaaaagaacgaaattaacaggactaaaacctgattgtcgaatcatttacc  
 45 aagtttgatgtcacgttgtaaaagcaggcttaaaaagatgactccttacaaggagtga  
 ggtggacctgggtgggacaggctagacatggccctgaaaaccttcttggggtgacaaagaa  
 acagactactggactgaagccacagcttccaagaaacaagaaaatgtagtggccaccaca  
 ttgggctttgttcccttatgagacattttccacctcatctcgggatcttactgttaccct  
 tgcccaaactgcttatggcatgaggggtccagagcccagcgcccagccaagtgtacaaa  
 50 agacgtttcctgtagagtgtgctgtgagggacaagcttgaggagtctgtagagcgtcc  
 agacaagctcacatttccctcattcatggatgatgaaggggatgtcacagcagaccagaa  
 actcctcaatgtctcaggaaaggaccgttttccagagcggcttacaagtgggactttctg  
 gggttccatctggagtgttggttttctgcttgccctcagactgagatagaagagcagtga  
 gacagaaagtagacagagaatgagctagcctccg

>scr\_gb-aw533305\_2 (TOXMARKER Assignment: 81; SEQ ID NO: 81)  
 ttttttttttttttagagggttaaagggtttcatttgccaaccggacagcctgagttg  
 gatcccctgagccccatgggtggaaggaaaggattggctcctgtaagttgtccactcttc  
 tgaagtatgtgactgcggtgtgtactgcccacatacacaaacaggctaggtagagaag  
 55 aaaagggaaccattaatagtcaacactgatacttatcaaaaatggcactagatggtgat  
 gggttaaaagcttcaacttagaagccaacagtgcagcagagacagacctctgttaaccat  
 60 tgcaggcaaaactgaaagacatgctcacacaggaagcaagcacaggcggctttgttgacg

gcttagctgaaacagactcaagacaaagcgtgttaacagacagacgcacttcacgggtgac  
 acgagggggccagctaccaagaagacattgacccccaaaacatgtatacacgccaacagaat  
 cccaaaaggcacagtgagaaaggacagaaggaaagttcgaaatagaactttgtgctgccg  
 5 aggtaggagattaacttccctggagatttccacagtggccaaaacttcggtgaggatat  
 ggaagacggaggtaccatctgagcttgatcagactctctaagggtgtgatattgcaaatag  
 tgcaagccaaacgactcagcgggcacatcacaggttcaagaccagcctgagaaacttagc  
 agggccctgtctcaaaattaaaagaggttggttttaaggaccagcctcgtg

>scr\_gb-aw915573\_2 (TOXMARKER Assignment: 82; SEQ ID NO: 82)  
 10 cacgagggatcaccagatgctgccaggtgctggttccaaggttgaaatgagaagtttct  
 gttaaactgggtacagagtttcagttttacaaggtgaagaagttgggcagacagatggtag  
 ggatgggtcacacaaagatatgaatgtatttactgccactgaagcaacactaagggtggtta  
 atctgagaagttatgtttattatttaaaggactaaattgtcaagctaacttcaataactg  
 15 ttttattttgtacaactgacatattcatatagatgacatctctaagatgtctttatcag  
 tatttaaaactgtgttacatctcttaggaatttgacacacagtttcacttgaaggccag  
 ataaccaattgtaggggtgcgttattaccagaatgtggtgggtccaagagcttgaactca  
 cgatcaagtttgggtgacacttgccctttaccactgagccttctcatcaaccaagtttcc  
 caggaattaagtaactctgtttccctaattccccttaagcaaacatggcagtcaagtgta  
 20 agcaggagacaggttatgatttgcatgatgaatttaataatgtaaccatctttggggaa  
 tctaattagtaccacaaaagagaaaaaaacccaacaggaaacagctgtctctcacaca  
 gtgttgagagctttccctcccactcattgccaatcagtgctcgtggtgtccctcacctg  
 cctctgtctctgcaacctgccagcctccaactgaacagacttccattcctgtgcaatcta  
 agtcagctctctccagctctctcctccctccctcctcgtctcctctctctataaagg  
 25 aaagaaagcactcactgggtataattgatgtctatatgcaggtgagggcaggtacaagat  
 aaggcaagacctgtgattgggcagtgaaaaaagaaaggcgggggagaggttttgtaagg  
 caggagagatgaggaggtagaagaaccaagaaaaaggcagagaaggacgaccagatctg  
 cgtggcttttaaccgggcaaggtagctatgaatatttcataagggaacagatttatatagg  
 acaatttgtcttacctaggtgggcagtttacatcaataccaattgggtgtgactttattg  
 30 tgtggacgttttgtggactgagaatttgctgatatgaatctgactgctaaattacaagct  
 ttgggttttgattttaactggctactgggagttgtgactgtagccacaggggcagatgct  
 gggattgtgagcaggggttcacagcacagctcccaggatggcagctgctgctgggccagag  
 aggagccagtgccaacatggggctagccatggaggtggagagatcgctggggacagagaa  
 gagcaggaggcagtggtgcttgggtgctggtgccccaccccccctgcatttttaatt  
 35 tattactgactcaactgggtgcttggcttttagtttcagaggggttagtccattagcatcc  
 tgaggagaagcatgcaggcaggcagacaggcatggtgctagaagggttagctgagagcttt  
 aaatcgtgatccgcacgctgcagagagagaaaaaggaaacagagatggagggatgactgt  
 ccttggaaggactttcaaaccttaaaagccacctctagggacacacctcttccaacaag  
 gccacacccctactccttcccaacagtcaccaactgtgaacaaagcatccaaatgtatg  
 40 ggccgatggggccatttctattcaagccacctcactgaaggaataaattaacatgtccca  
 aagtattaaatgtagtcatttttctcagtactgagacaaaatatctcaagaaataaaaaa  
 aactgaaggacgtatttcgtttggctccccctttaaagaaacagtcaccatggccgg  
 gaaggcatgtggctgggtcagagtgccctcatgcaggaagcagagagtggggagatgct  
 45 cctcgaagccttttcccttttatttagcatgcacccaagcccacaggagctgggttaacc  
 cagcaagccttgcctggcctggaagccaccccaacaaccatcatcaccacagtgccctcc  
 tacagtggggattatgagttgccaccatgctgtttttcacatgggtgcaggggatttgaa  
 accacagcctcctgctgtacagaaagcatcctgaggagccatctctctggattcacccct  
 tcacttttggctgactgggcctgagctggagtcacctgggct

>scr\_gb-be108509\_1 (TOXMARKER Assignment: 83; SEQ ID NO: 83)  
 50 ttttttttttttttttgggtttgttttgttttgttttcttttcttaaggatagtt  
 taaaatacaaaataaagtatgtgatatgtcaacatgatcatgccccccagacac  
 agcctttaactgtccagctcaaataagagaaatgtgaagcttaagatgtctttgtcctc  
 aggaagacatcacatgtgtggttgctgacactgcacatggcagcttccccacaacatg  
 55 ggcccttcgccttcacactgacaagaagtgtatgcccttcacactgacaagaactgtgtg  
 ctactacaacttgattggttgaccttccccaaaagcagtaattgtatttctcaagatg  
 tctaataatcaagtggagactctcctctggaaggaaactggactcagcctcgtgccgaattc  
 tt

>scr\_gb-be111483\_1 (TOXMARKER Assignment: 84; SEQ ID NO: 84)  
 60 ttttttttttttttttcagacaaggatgggtttattgaatggacccccctgagactgatca  
 atcaggggccagggccgcagcctcagaattcaggggctgagccatgactctgaccatttct

5 cagggccggcttataaagggaacccacaaagccacaatgagctcgcatgcaggtgct  
 gccggatgggttgctctgactcaagccatttcagacagaacagctcatatttacctttaa  
 tgtggtgggccatatgtaaagctttgtgtaatttattaagttgaacaaacctcacagcat  
 gaccttgctctgagtcgagtcattttctgtatcaatgatggcagggcatggaacaaatgg  
 ctatagctatgctaggtgggtagacctcaacaggataagaaactaaaaagtaacaaaga  
 tgagaagacaattgggcatcctggt

10 >scr\_gb-be120910\_1 (TOXMARKER Assignment: 85; SEQ ID NO: 85)  
 tcttttttttcggagctggggaccgaacccagggccttgcgcttcctaggcaagcgctct  
 accactgagctaaatccccagccccaggaacaagccttcttaaacaccaccccatctct  
 ccagtcctgatcaatattttatgactacgtttactctgtaaaaacaaaggattaaaatct  
 aatccgattaccagtccttagacaaaccttcctcaatctgagttttctcaagtataaac  
 acttcacaacaccttctgagaaatgtccacatcactcaaagacaacacatttgggaggtt  
 15 tttatgggcttctttcctacagaacctttcaaagcttgtaaaacttcgaacctaggggac  
 atttgggagttcttctcggtctcacacaaaacggacttgctttcaaagatcccttcggat  
 tctatttgacttagcaaaaacacagcgcgaaaacacaccctgtaagaacaaagggtgcaat  
 tg

20 >scr\_gb-bf285287\_1 (TOXMARKER Assignment: 86; SEQ ID NO: 86)  
 agttcatgtgcattggtgtttgtctcacatgcatgtctgtgacgtatgcctgtaggagggc  
 atcagatccctgcaactggagttattgacagttgtgagctaccatgtgggctgtgggaat  
 taaacctcgaaaagcagccagtgctcttaaccactgagccatctttccagccacctcaac  
 tcattcttaaatccacttaagacatagaggaaacactattccttctattctgtttgctga  
 25 tatctgtaaaaagtagacagacttgcagagtggtggtggcagacacctttaatgtcagcac  
 tcaagaggcagaggaagacagatctgagttcaaggctaccctgatctacagatagagttt  
 caggtcagtcagagctttatagagagacctgtctcaaaatacaaaaaacaaaaccaaatt  
 aagtagacagactccacttacacgaaacgtaaacactgtttcacacacttcagaatcac  
 atttaaacatacaacaagaactgacagaaccaatatcaggaaacctcatccatata  
 30 aagcaacgtcacagcaccaagcagttaacagcttttggctcgctctaatacgaggatccca  
 aacacaaatcttacacagacatggggaggtacatcctacatctcatctcggtcgcagctc  
 atcgctcagtcctagggatcttttgggtccccacaaagatggaggcatagccttgctcttc  
 ttgcccagacaaggaggccagcagggccaggaagttaaactgccaatacctgccaatgctgg  
 tctc

35 >scr\_gb-bf390383\_1 (TOXMARKER Assignment: 87; SEQ ID NO: 87)  
 ttttttttttttttttanagnagctgtattttctttatattctgcatgggatatgaaat  
 aggggttttgcctcacaggagcctggtcaatatagacaggatgtantcagggtgtgtct  
 tccaaggatcatctccatttccaggcagatggaaaaaaatcatgaacaatcatgttgatg  
 40 attttgaaagatgagtataggcaatagcatgtgtcctctgtcctgagcaacagatctcag  
 ggatgtgaggggtgtgcgctttctggatggttcaccatacgcactcttcagcaccaaggcta  
 tgcaagctttgttcagtaaggcagaacatcaggaactcaggagagtgggtcccggaagggt  
 gatcatgtggcttgacccctgattatccatcttctcaccaatgggttgcttacattcga  
 agcttaaagccttaaagtttaacttcgtcttgatgctgttaaattgttttcaattacagc  
 45 acgatc

50 >scr\_gb-bf558463\_2 (TOXMARKER Assignment: 88; SEQ ID NO: 88)  
 ttttttttttttttttaagagaaacatttttaatatctgcaggctcacgcaggattcaa  
 ctgtgtgtggtacagtcctagagtgacttgcttctatttacttccacacacgggtgacttcc  
 gatgagatgggttaagctgagcagtatatacttctgaacagtgccaggatcctgttttca  
 55 aacagctttatcaatcgaaacatcctcaaagagccattggaggcagtggtggctgggcat  
 ctgcactaaaatcgcttattcagaagggtgtcaaagcagccgagggccttgagccacaggt  
 tgctgggtgttcacatctcagctgggacgtgataaagactgcatgagctgcagatccgcaa  
 acagccttgaggctggctctgctcctgcaaagtcaatggagccacaagggtacttcttaa  
 tgggtgtcatctgttcagggtcttccaggaggttaagggaagcactgtctttgcacacagtc  
 60 tctatcacaagggtcttggttagcagcatgagagttccctctcagccagggtgccacagt  
 gagccatctattgtcctcactgcagagtgcacaggatgaagatgtccactttcctcatca  
 gactgtgtgacagcctcatttctgtgccaaacggatcagaccacactttcaaccctgggtg  
 ctgcacatcttctggacgataaccagctcgatttacagcctgctccttctgggtattcttc  
 cagccgcagaaggggcccgaagtagatagggtagaaggcggctccgaccatagagatgaa  
 gcctccgaatatgagcgcgggtgcgcaggttccggggccgcccaggtggtgagaaagggggc  
 tgcaggggcgggcgaaggccccggcacgctccgaaacccgactcccagccttaagggtcgcga

cccggtcgggaagaggcggag

5 >scr\_gb-bf560709\_1 (TOXMARKER Assignment: 89; SEQ ID NO: 89)  
ttttttttttttttttttacatttatgaatttttaataactcctgtcaagatcttacaagga  
gaaattacttttgggaggtgggtatggaggttagaggtaggttggaaagtggatcatgatc  
tcaaaatagtaaatgctagctgagtggctttccagagagaagcgacatgccctgacgag  
actggagaacatgtgtaaaggagagcttattttcaggtctccgctggcctccatcctctt  
caaaaacctcagctcctgggttctgctcaaccacattctgtaataacttgcacagtagg  
cctgtagcaccttgtaggatacagacagttccaaatggatgtccactccagctctctggct  
10 gctctattctgtacttctcttgaatcacagcttttatccatgtaagtagatgcctttacc  
tgggcacttgaaagttcagaggagacaggtcttttagatagaaatgtgcaaattacttatgt  
ggttattgacaatcaatgactgttctcccgtagtctccctcgtg

15 >scr\_gb-bg662990\_2 (TOXMARKER Assignment: 90; SEQ ID NO: 90)  
ttttttttttttttttttacataaaactattttattttaataaaaaccaggactgacctct  
cccacacgcaccagcacatgcactcgcaaatcatgtcctccgtttctgttcctcctgaa  
cagccacctcaaaccacacaggttttcttctgtgacctccttgaacctgaaaattggga  
gatcccatgcgaaacactggcactcttcccccacacctgggcaagcattctcctcatcct  
cctgggtgggacagagctcagctcttccaaaggcaccacagatctgggtgtggtttcccttca  
20 cacaaccgggaacaccaatacccagagctgctctttgaggctgggacctcgttccag  
gtcaactcctctcacacaacagaggaggtttgttaacctgcttaagcgctctccaaagg  
ttcctggcataggtaccgtctgggtatgaggaagagcgacagagagcaattgagcaccaag  
ttccctaagtccacctgaaggaggtgccaagctccagttcagctctgtaccaagaaaaa  
gcaagcctagcgccacacatggggaaggtggggatggcaaggtctcagccttgagaatct  
25 cacatctctaccctccagcatagatcccatgagggacctactagcaccttggcgattgta  
agggctcagcccaactggagacacaccacacaaacagtgggcatttggagttggccaaa  
tgccctgtgtcggttaacaggggttgactcccgcatctaacactgactgaaggacacacagc  
acagcagctaaggtcacgagaggtgactgacagaaggtgtgtcttccagaggcacatg  
gacatttcacacactgctcacaggcaagctgggacaggagaagagcacaggctgccaggg  
30 actcagcagcgtatctagggcagtcacctct

35 >scr\_gb-bi278552\_1 (TOXMARKER Assignment: 91; SEQ ID NO: 91)  
ttttttttttttttttttggggtttggtatcatttattttttttcttaaacccacttgta  
gtttgggttcagctgggaagcaggatatacgggtagaggggaaggggacgggtgcgagcagg  
attggcccatagctttgggggcaatctccaaacctgctccagggaggtaggtcctgttg  
tcaggctcccagctggctcaggtgaggctccaaatggatcttctggagcaagtgtcctga  
gcagaggagaatttccattctctccaacctcctcaaagacctcagtcagaaggtttt  
ccaacacagtgcacaggcagttgaggggacatcaggccacgggcaggcctgagtgggtggg  
acaaggaacactgtctggcttctggttccaggttaacaacctaggatgtggctaccagag  
40 gctgccatctagagtgcctccgggagctgcttctctgttctcctgggctgectgggatc  
caaacctgcagctgccctgggtgcaaccagtgggtatacttccacccccccacctcaga  
caaaataaaataaaaataaaatacaataaaaattagaataaaataccaatcgggtcaacatt  
tacatttacacaaatggacaagatgatcccccaaacctagaaagtttacagactggatgg  
gaaggatacgcagatgaagatgggtttggggaggaagaggttcgccgtgggtggtgatgg  
45 tggggggtcctggcctgtccaggggagggccagagccctgcaggaactgtggtctcaga  
gcttaggcaatacggccagttcatgaggagaacagtgcctgcaggccacttgagtagaa  
aacaaggaccaacttgtcctgacaggtaggggagcctaaaaagggtcaatatgagatcgc  
catggccagcaggacaccacagtttgggagaggtctcgccctcctgttcatccattcagag  
gcggctttgataggcgtccctctggcagcgggagagcctctggcctggggaggtcaggg  
50 tctgtgggtacctgcaacgcccacttcccctcgtgccg

55 >scr\_gb-bi278749\_1 (TOXMARKER Assignment: 92; SEQ ID NO: 92)  
tggctcagtggtcgcagcacagtaacaacatggagattctaaaaacagagaaagagaaaag  
caagaagatagtgaggaggagacaagaagaaggctggggggccagttttgtatttttgt  
tttggttcagctatatgctccacacttccaaagcagcaaatgtgttgcataccaccccaa  
acctgagaaaagctacagcatcactggcaaggacaagctagcgcacgggtgacatcctcta  
accctggcaattgtaattatacaactgcagtttccagcacacaccattgcctccgacact  
attggagagcccgtgacactccaaaaactgctaaggcctttacagtatctgaccttcaat  
ggccccgaaaactggtaggcgttctccccattccaacccaaaaattacatgcgagcaa  
60 cggaagagaaaagcttttaagcccgcgcgagcgaagagaccagcgagcgtgctgaagac  
cacagaccaggttaagccagctgaggctggagtttattgccgatgagcgtgagtcctggg

gaggagcggggaaggataaggtcgggcaggatcaggaccttggctaggagaggcggcgcc  
 acgaaggcgaaggcgggaggtgcagacagacaggcgcaggccacgggtggggcgggccag  
 gctatccaggcactcgggtgagcgggtctccggcgctcgctcccggagctgggtggcggtgt  
 5 gggggcggtccgcggcagtcctggctgcgggtcgtggccaccggaggccccaagcaagc  
 aggacgcggcgggaggcggggcggggtgggtgctgctcgagcacacggagcagctgcagcgc  
 tgggcaaggggtcggcggggcccgaggcgccgcgtggggaccagatgagcccgtagt  
 ataccgcaagcaacacagcagccaaggatacacacaggaagtaggcgcagacagggcgga  
 gccgcagccatcgggcgcggggccctcgctcagccccgtaccacctgggctctcgccac  
 10 cactgcccacgcagctcgagccccgcgtgctgcccactcagcctgtaccgacccgc  
 cccaccccgccgcttctagcaagccacgccccttctagagtacgcacctatcagaccgc  
 cccccctcgtgccgaa

>scr\_gb-bi295938\_1 (TOXMARKER Assignment: 93; SEQ ID NO: 93)  
 15 aaaaatctcgatgccctcaactgttaggttaaagcctgacctgtgtcactatgtgctgtg  
 acacgaacctaatcccaagtggacagggacacctgagtggttctcgtgcttcagtcc  
 ttccctcatgattcttgcgtgggtcctcttccactgaggtctctcccctgagtcataatatta  
 ctggaaaggctacctggagagcctttgaattgtggcattcctttttaatgtgtccctct  
 cttccacagatgaacacgcgcttttctcttgagtctctgtcatcctgtctcttccacttt  
 20 tcggctggtgtcctgacaagtttctccggcccagggtcaacagctgccctcattggcttg  
 gctttggcagctgtgcacggtgcagccttgtcttcttttgcgtgacacttcttttctgtg  
 tacttgttctgaatttctttgtcctctttgtcttcttttcttttgcgtctctgtgtacctt  
 tggtttggggatcttctcgtgtctcgccgcgcctcacttttctcctcatgggacagtcc  
 ttcattgaagtggccaattttccacagatccggcagcacctgtcatttggggccagttct  
 25 cctcagtcaggacatccggatcaaagaagtatgccaggatgtcctttggaaatcctttg  
 actggaattccaaatactcttctaccattgataaaaagctttcattataaaaatttgtcatt  
 ttcttgataaatccagcaccagattgtggttcaaatacaaaggatcttcaatgacgatg  
 tattttgaggtccactgtttcttaaaagttgtaagcagacttttctctgatgctgatt  
 acgtgttcttaaaagtcaaaactcctcagtgtagaagcgtagaagtccaaccacagctgc  
 30 ccaacagattctgtattttttccatattctggccaacaagtgggcagttcatttatattga  
 tcgaaaaagtagatattccagccatcaacaagtatttctgggttcttttcacctttgtat  
 atctcctgaagcacagggatgacagggggggaccgctgctggaggaagtacagcaccata  
 agagtgtaaagcgtatgatgacaagctgcctctggacgcgtcaccgatgtcacacatcttt  
 gtgaacactttcatggtgtagcacaggtatttctcctggttcaatggctgagtatgca  
 35 aacaggagccgcgtgttgtgaagagccagtggtcctcgtg

>scr\_gb-bi296376\_1 (TOXMARKER Assignment: 94; SEQ ID NO: 94)  
 tttttttttttttttttccaggagtccttccggtccctgatagcgggagcctggacctc  
 tgaggccgagaggggtgctgtgtccccggcctccgagccgaggtggcccggtaggggcg  
 40 ccacggagtttttttttttttttcttttcttttccaggagtccttccggtcccagccag  
 cgggaccatagacacttttgaggccgagaggggtgctgtgtccccggcctccgagccgagg  
 tggcccgctaggtggcgccacggatttttttcttttccaggagtccttccggtccc  
 tgatagcgggagcctggacctctgaggccgagaggggtgctgtgtccccggcctccgagcc  
 45 gaggtggccgggtaggggggcgccctccgaggctttatttttccaggatcctccccggtc  
 cctgccagcgggagcatggacttctgaggccgaggggaagctgtgttccaggctatctac  
 catggcctcctcgtctgtgagcactcagggttctaaggctgaccagttgttcccttgcg  
 gtccggttctcttctacatggggacctcttggggacacgtcaccgaacatgacttccag  
 acgttccgtgtggcctgtcatgtttatccctgtgtcttttacacttttcatctttgctat  
 ctgtccttattgtacctggagatatatgctgacacgctgtccttttgactctttttgtca  
 50 ttaaaggacgttgaagaggcttgcaccaaggctgtttgcttgtccagccctagctctt  
 tcttctgcgcagtgggcctctctcgatgcttgaagcttagcgtcccccatgagtacgcgt  
 tctgtcttcccggtgcttgcctgtgctctgtggggcagctttatgacaaccgtccc  
 gcgtgtcaggcgttcccgatttcccgtgggtggtgtcgtccgttaccggtaggagtcgt  
 tgggtgccgagtgcgactgaaaggggtttcccgtttgggtgctagtacccccctggcgtgct  
 55 cctctgcggccgacgggtttttttatttgtttttttttttttttgtttttttttgttttt  
 ttttttgttttttgaaggagttcccgaacctccgctgcttgggtggtgtgtccctttctt  
 tctgtgtgtgtcctcccaggttgcaccttttctccttcgaaggggattttattttttta  
 tttttattttttttttttttttttttttttttttgaaggagttcccgaacctccgctgcgt  
 tgagtcctcgttcttccacgccacgtgcctcccaggtgcaacgcttcccttttttctcgc  
 60 cctcgagaagggtaaattttttttttgggtgtgtgtgtgtggcagtggttagcgacttcttcc  
 cgtgctctctctcgtctctctcgtctattcccgtccagtgctgttagaaagctctca  
 cgcccgttgttcccgatgcatggcggtgtctcgtctcccgttggatcgatgtgggtgctgcg

5 cgttctcttcgggcccgggccaagccgcgccaggcgagggacggacattcatggcgaat  
 ggtcattcagcgcgaatggcgaccgctcttctcgttctgccagcggggccctcgtctctc  
 ctccccattcctttgcaggggtggtgtgtggaagtcaggggtgaggctgtccggcacgagc  
 gctgacccgcgcacacttgctgctgtggttcgcgggtgctcctgtggacgtgtcggggcg  
 10 cttgccccacgcggttactgcttcgcggccctcttcccccgctgccgggggaagtggt  
 agaccgcgtgcgggtgcatacccttcccgaaatggtgtgtgcacgcgcctgctttgtgtga  
 gccttgcggtgctcctggagcgttcggggctttgaccaccaaggtgcccgcttctgagtt  
 ggcggtggcgcttcccgcctcccggcgctcctcctgtgctccatggtgcttgtgccttta  
 cgctttcccttgctcctagttgcccggctttctgcacgggtgacagaaaggggggggtcgag  
 15 gagttagtggtgcgggttaaaaggctccttccgttgggtgagcgccaccctgcttatg  
 tttttggtgcttcacccgcgggcccctgcgcgggttaggtggtgctgagcgatcgcggt  
 ggccctttttaagaccggactcctcaagtcaaggctcctcctttgtgtgcgccttgaa  
 gaggcctggccctcgggcgggacctgtcgcagggtcccccggtccgcgaatgctcaagaa  
 gaccccgagagaaagagacctttgccgataccgcagacccccaccagctggcgctggtc  
 20 cttcccgcttctgtcccgcgcctgttgctcgtttcccggtgctgcaggagcccttggt  
 gctcgtcggtgttggttctgctcccgcctcagtgaagaaattgccttctctagctatctt  
 cgaaagggtcttacgatctccgaggggttctcccggatgggtcccctcggtgcgcccc  
 ctgacctcagccttctgcgcgcagcgtttgctctctgcctaccgcgacccgcgcctccc  
 cgctccgagtacgaggaggatcacgcgggacggggctctgtcgacctgccgctgtgcgg  
 agcttgtgggggagattgggtttctggtggcagggtggcggggaagggccgtgcac

>scr\_gb-bm384392\_1 (TOXMARKER Assignment: 95; SEQ ID NO: 95)  
 25 tttttttttttgtttttcaagttgcacattttaatttacaatgtttaccagtaaaaagg  
 attagttacaaaaaggaaagctgtctgtacaaaaataagggtttttttttttcacattcat  
 aaagagaacccactgtgaattcttaccttgtagagtcactcaaacagctcactttgg  
 taaaactatcttggaaggactagtaattccaggcaagataataaaattatcagcttccaa  
 30 tcatgtccaggagaaagaattttctgaacattttccctgtacagaaaagctctctgtact  
 tgcagatccttagaaaagccagtgctctcaggagacagcctggtaccaggacgaagcata  
 atctcctgctcactcaaatggcaatccttccctgaatctgacagacacacatttatcatag  
 cctcaggtcagcaggagaaccagatggttcaggatcagcctctctccactcaatagttta  
 tcatataaattaaatatggagaggtacacatgagaaaagggggagctctttttcaaactcc  
 cacttcctaataataacacatcacagttttaatgagcagagaagggttaagtaccctgg  
 35 tttgggcacatttccctcaagggaacaaagatcaaaagccttcaaagcatactggc  
 ccgtcccactgcagccagcctgattccagaatgaaagcatacagtagctgtaaaggc  
 ctggagccttcagaaaagctttatatttagtgataagctgagctctgctggcaaaagcccacc  
 tataaaaaggagcaggtctgattcaciaagtgatatacatgcatgacccaaggtaatgaa  
 gaccttcaaatgcaaatgatcctaaagctattggaacctctaattacgagtgaccggttc  
 40 agatgtgctccattagccttaaaaactgaccaacacacatctgaagaggcacttccctt  
 agcatataacataaacacttgaccagaaaaggcatgggtccaaaaaacagttaactaaaaat  
 ttagagtctaaacctctcttctccaccgactgaatgaacacaccgcgaatgaggaccaa  
 cagaatcagtgctccaggagcgtgtgtctgtctggccatgtgatcaggaacctcttaac  
 atagcacagcacagcacagctgctctgggcacacaaaagccagttcaccatgaagaaac  
 45 acaagggattgtgattaaacctatccctgtgtcaggagcaactccactatggttttgat  
 cactcagctcagagggataggagtgcctagcaacaagtcctaactcctcgttactccagt  
 ccgggcccctcactgactcagagggtgcctttgtgtataaatatgtgagaggcagcaaatgg  
 cagcactgctgacagggtaatgcaggccccacagcggagaaagtcttctcctctgctgctc  
 caatcttctccctacagttacagtcctgccagtgatggccaaggaccatgtgtgagccag  
 50 ctctttgtgaccaagctttggcaagtcagtaagtttgtcaaaggcaaaatccttctgtgg  
 acaatgctagctgcagctctggggacgtgtgagagaggagagggtcctctgacgggattg  
 gggacgtgtgagagaggagagggtcctctgagaggatttgactcatcagccctcttgcc  
 cagttcattaatcagaaggaaggggagaggagaagacagcagaacatgagtcagttgtga  
 aatctgcacagctgacatttgctcttcacagcagaaaggacttgaatgagaatcatgaaa  
 55 cttgaggaacacttgattttccttcgggatttaaaaatgtgtcttgtacaaaagacta  
 cattcagtggtgaggtccaagagcggcagcaagagctcggccattaaagcgtgccag  
 cactgggaggagactgtcatctgcttagcatgggtggtgagcaggccagggtgctcctc  
 actggtctccaagtcggaagccctggccccagttgtgtctcccacctccgccattctgat  
 cagcagctcgctcatgcttgacgggggcacaccgaagcccgacacccctcctctctgc  
 60 tgggtagccagcgggtacaaaactgaggtgtggacagaaaattccttctcccaaatcca  
 ttgggtatctgaacatcaggaagaaataaagatgtccgacaaggtttccaatgagctcat  
 tgatgaccgagcctccaatgatatagttgaatccgaggataacccaaggttaagtaacagg  
 ccttaaatcgtgttccaaacaaaatgatataatcaggtctctgttcagctgggcccaga

cgtaaagtactgacatgattagaggaatcatcagcaactgcatatccatggctaagccag  
taataacaatgcagatccagttg

5 >scr\_gb-bm387477\_1 (TOXMARKER Assignment: 96; SEQ ID NO: 96)  
aaatttcaagaggtcagagtggggccttagattaagtaactaatgcacagcaaacgctgt  
gagattaggtgtgaaggagctggctgccctcctgtctcttcccttctctatcccacagga  
gctacagagagagcacagcagccagacgctggcacaacaggggaacactctttatgccaa  
tcgcaaagatgacaagcggcatgaggaggacccagggccctcctttgtgtggaaggacgg  
agaggttctgggagggctgggaagggatgggaggatcctttgtgtgggaggattgagga  
10 aggcctgggaggtcctgggaagggctaggaccgctctcctttgtgttagaggtctgggaaa  
gtctgggaggatcctcctttgtgtgggaggactgaggggctctgggagggctgggagggc  
cctcctttgtctcacagtttttagatgttgttccatctgtctcggagtttgaatttctgg  
atctttcctgagacagtgagaggatagccttcacaaacacgatgtatcggggaatctta  
15 aaatgggaaatctttcctttgcagaaagctttgatctcctcctcctggtggtctctccg  
cctcgtgcc

>scr\_gb-bm986259\_1 (TOXMARKER Assignment: 97; SEQ ID NO: 97)  
gtaacccacctccattctgttcttcggacgcttgccagtggtcaattttatcttctt  
tcaaaaaataaaagtcgagtgcatccagagacggccttaaggcaatacgcctcatcttccc  
20 acagtaaagatggcgacgccgtgagtaagttacaagtaactccacttccgcaattttctt  
gagccctgggtccaagatggcggacgagggccaccggcggtcgtgtctgagatcccgggtg  
ctgaagactaacgcggacccccgagatcgggaattgtgggtgcagcgactaaaggaggaa  
tatcagtcctctatccggtatgtcgaaaacaacaagaatgcccgaatgattggttccga  
ctggagtcacaagaagggacccgggtggtttggaaaatgctggtacatccacgacttc  
25 ctcaaatacagatttgacatcgagtttgaaattcctatcacatatccactactgtcca  
gaaattgcagtccttgagctggatgggaaaaacggcaagatgtacaggggtggcaaaata  
tgtctaactgatcatcttcaaacctttgtgggccaggaatgtgccaagtttgactagct  
cacctcatggccctggggctgggtccttggtggcagtggaagtccctgatctgattcag  
aaggggtgtgatccagcacaaagaaaaatgcaaccaatgaaggatgaagcttctgaggcag  
30 gacagagggactgttgctagactctgattctgtttcctccttctcatgattccttcaag  
ggtcacctctggccattacaaagtagctggagggaacaaataacaaaacccaacaaaagg  
caaggtcacaaagttgctaaattaagctgtacagagaggtgaaagatttgggccttga  
gaggcgggttgatcccttctccaagcagagccctggaggcattttggagacctggggtg  
35 taactgacagcatatagctttttgtttctggagacaacctgtcaataaaagctgcttcc  
catgggtgtgaaaaaaaaaaaaaaaaaaaa

>scr\_gb-s69874\_5 (TOXMARKER Assignment: 98; SEQ ID NO: 98)  
tttgctatctgcacagcccatcgaggacctgaggtggcaaacctggacagtggtcag  
gcggcgctcacgtctggggtgacaggatgaagcgggctgtgggctgtgtggagcacctg  
40 cacccttagcacctttgggtttcttgtggagttctcgccccagacatcagtgactggat  
tgcaaaaggcaattcatcttttattggatcaggagcgccatttggaagtgtgccattatgg  
gaggtcgtagctgtctgtccctcgtgccgaattcggcagagccccctttttttttt  
tt  
45 ccaaaggaatgatcctgggttaaaccaacggtctgacatgggtttcgggtaaagtgtctat  
gatgaaaagtcatgaaaaataaaaccaaagaagtgaagcagtggttctgtacgacctg  
ctcattgaattgagcttattccctcagccagctgactgctgtccaggatgacgagttagc  
cagtcctcatgtaccttctcatagaccgagtagacatggcattgttcatgacgcactc  
caccaccatcttcccgctcctcagttttctcgttatcgtgctttcttcccttccactt  
ctggtgctggaccagggaacgctctgtgaaggtgcagaccgtctcagtttctcctgccatc  
50 agctgtggtttcatcaaacttctctcccaaggtgcaagaaaacacggctcgtcttcacgt  
gctctcagttttgacggtgaggttgttgccgtcgagggtaatgatgcagctcgggttggc  
catggcaccatcttccaaagaccagccctactcctagttccttcatgtagtcctcaaa  
cccggtggctttccaccagacgccacttcccttccaggtccttaaggctggccatggcgag  
cgggagagcacaaaagcagcaaggagacgcggtggcggggcgctgaggggaataagctca  
55 attcaatgagcaggtcgtacagaaccacactgcttcaacttcttgggttttatttttcatg  
acttttcatcatagacactttacccgaaacccatgtcagaccgttgggttaccaggatc  
attcctttggttagtaataaatgcgtttgtgctaaaaaaaaaaaaaaaaaaaa

60 >scr\_sc-119263563\_1 (TOXMARKER Assignment: 99; SEQ ID NO: 99)  
ncctagcagaacgcttgtaggagctgtgtgggacaagatagcctctgataaaataaactc  
taaactgaactccttcaagaaaaaggactggactccaccactgttcaataaagtcacag

cgaggggatgctagagggcggtagacagaaattaagacattctagatacggggagtgggccac  
 ttggttggggccaccacttgcccttagcataggtaccataggttaagcatggaaggcagtaa  
 ggggtggatgtcatttttaagttagagagcagcaaatttagtacatggtttatcaataaaagg  
 aaaggagtcgaagatcaatctgacaaatagatctatcagctgaattgtaattcttgggggtg  
 gaggggtcagaggtccggcaattg

>scr\_sc-132556005\_1 (TOXMARKER Assignment: 100; SEQ ID NO: 100)  
 caattgctgctctaggatagtcagagtggttctctgtctcctgggaaacagtggaaccag  
 gaatgaaagcttcaacctgggtaccagatttttagatgttttagggacaatcagtcgaatt  
 tttgtgtgaatgtatgggttttatatgactataactgtgtaagacagagaaatggatgtac  
 a

>scr\_sc-132570828\_1 (TOXMARKER Assignment: 101; SEQ ID NO: 101)  
 ccatggacataactacctcctgattaagtcggttaattgagacctaatacagtcctgttaga  
 ttattgaaacaggtcctgttagcagactgcaggagaaaaacacgggtcatgaaccaagag  
 tgagtcggga

>scr\_sc-132947646\_1 (TOXMARKER Assignment: 102; SEQ ID NO: 102)  
 aagcttccctccatttccagtagtgccatacgtggcaaccataggatcc

>scr\_sc-133387221\_1 (TOXMARKER Assignment: 103; SEQ ID NO: 103)  
 aagcttcaactgtctatttattcacagtcacactggctgagatgtcctacactgtgtcca  
 gtgcaagtgtgacactggacattgatgtcttcttctgtatcttagaggaaaggctcggt  
 gaggttagagcctggcttccggcttgtcatatgacccctaagtgattatttctactgta  
 ccttattctcagaggaattttatcatgaaaggggtccaggagtcctcccacaaaccttag  
 gaacaccaatctcagtcagacagggatgtttgtaatgcacacctaagtcctgatca

>scr\_sc-133555783\_1 (TOXMARKER Assignment: 104; SEQ ID NO: 104)  
 gctagccatttgggtattttattagataacaagtttagggaactcatgccttggaaagggtgt  
 gttgggtgctttagtatttcttctgctggcacagggaagctacagctattatctcaataaaa  
 tagctgtcccttggaatttttttttttaataattgcttattcgagccaacatctaaata  
 aggtgcatgcattgtatttgccttgatacgttgggtgtgtctcttttcttcttctgtgaag  
 tttcttccctccttatttttttcttctcctgtattgtatttactggaaaaaccagatcgcg  
 cgccctgcaggttctgtaca

>scr\_sc-133678871\_1 (TOXMARKER Assignment: 105; SEQ ID NO: 105)  
 agatctgaaagttaggcaaaatataagagcagccctctgaagaggggacctgccagctca  
 cttgggactcaacattctactgtagagctagc

>scr\_sc-133725675\_1 (TOXMARKER Assignment: x106; SEQ ID NO: 106)  
 agatcttgggggttccaggcttgggtggcattcaattttaccttctgagcccaggagcgag  
 aatcttgaactaaagagggcttgacagtgctagc

>scr\_sc-133955481\_1 (TOXMARKER Assignment: 107; SEQ ID NO: 107)  
 caattgaacagtagtctgtaagtagtgcaacactgtaaaatgttctcttttagttcagaga  
 gaaaattcccaagcattattccaactgctgctaaaatagatgtttataattatcagtttaa  
 tgccagttccaaacccctaataaagcaaatattactgttattgccagcaacttctgaaa  
 ctacacaaattcagtgatccctccctccctcttttctttcagtcagtgaggagcaga  
 tacaaccagggtccaagataggttaagtgtccttagatgatttttagatagcaggtgggtg  
 caaacttttaatcccagcacttgggaggttaaacaggtggatcc

>scr\_sc-134521597\_1 (TOXMARKER Assignment: 108; SEQ ID NO: 108)  
 ntaacaaagatgggttagagatccaggtcaccaatcctcttctcagacagaccatttc  
 tggggtcaacagccattactgcatgtagagtaaaagggaagtaagacagagagagttcatg  
 ggagtcctaactggctgtgtggaacagctttccaattgttctgggaatgaatgtagag  
 tcagtgccctgcatgggtcatgataagagtgcctgcaagtgaggcgctcacaagctt

>scr\_sc-172126480\_1 (TOXMARKER Assignment: 109; SEQ ID NO: 109)  
 ctcagggtggccttaaaactcactatataactcaaggatgaggttgaacctatcttctctatc  
 tctgtctcctgagtgactgggattgtacacatgtgccaccatacctggcttacgtgatg  
 ttgtggatcaaaccatggctttatgtatgtctaagcaagcactttatcaactcaaccaca



attcatctctatatattttaaatgtaatatctctaatatgtctttacattttccagctacat  
tcctagg

5 >scr\_sc-172130231\_1 (TOXMARKER Assignment: 110; SEQ ID NO: 110)  
tgatcaagagtcacaaacccagagagtctggggtgctgacatctgaatgtggctggcctg  
ccctggctgactgctttcagtgccagccacactgatgcccttagccctctggggttaat  
ttaggaacttgaggctcaggccaccgtcaccagcaatgaactcacaagaatgagatgtgg  
ctgttgatttcctagg

10 >scr\_sc-172755010\_1 (TOXMARKER Assignment: 111; SEQ ID NO: 111)  
agatcttccggagcaatgggggttcagcttttgcagcgcctactggacacgggagagactg  
acctcatgctggcagccctgcgacactggcggcatttgctctgagcaccagtctcgga  
cagtgggcagccctgagtgctcctaggaactcggagagtcgtctccatcctggggtgtggaaa  
accaggctgtgtcgctggcagcctgccacctgtgcagggttatgtttgatgccctcaagg  
15 aagggtgtcaagaaaggcttccgaggcaaagaagggtgccattatcgtggatcctgccggg  
agctgaagggttctcatcagtaacctcttgaggcttctgactgagatgggggtctctggcc  
aaggccgggacaatgccctgacctcctcattaaaatggtacctcggaagtcaccgaaag  
atcccaacaacagcctcacactctgggtcattgatca

20 >scr\_sc-188295137\_1 (TOXMARKER Assignment: 112; SEQ ID NO: 112)  
gctagcttaagggttcttctgtaggccgcctcatttcctgggttaattttactttatgta  
tatgatgttgcttgatgtagatct

25 >scr\_sc-190079504\_1 (TOXMARKER Assignment: 113; SEQ ID NO: 113)  
agatctttttgcttcccttcttttattgatccttaggaataaatcctccaaactctg  
ttgttttttaagttttttgaaagacctgatttttttccattttctttgcccttgcaaat  
aaccatcagtgtaattagttgtccatgctgcaagggaataactttgtgagggaaataagca  
agaattgagtggtgtttactaagagggtcacgcggatgggttttgggtaattatttactag  
30 t

35 >scr\_sc-191455923\_1 (TOXMARKER Assignment: 114; SEQ ID NO: 114)  
tccggagctggggactgaacccagggccttgcttccctaggcaagcgtctaccactga  
gctaaatcccccaaccccgctcaaaggccatttttatcctcatcaacaattataccttact  
ttttgagttggaaatgtaattcagtaaatagtcgtgttttccctagtagtacaaaagtcttg  
gctccctcactaaacaccaaaggaaagggaagagctcacttctttgactttcagtg  
gccttccactcagactatgcttgttttagaacttcggcagcttttttcatgctctcctcca  
tcttgaactcaacaacactataaaaaagaaaagccaaaaacaaatgaataaaaccagtct  
tacttggaattgaacttggaattt

40 >scr\_sc-195460151\_1 (TOXMARKER Assignment: 115; SEQ ID NO: 115)  
tctagagaaatatacatagacagcaaggctggagttgagccaggcaacctaagctgggcc  
accggagtcaggcagctgcagaaggctcagtgagcaggccagtgctagcctgtgacgga  
gtgatgtagacactcagccacaccagggaagccaatctccaagttgtcttggttagactgt  
ggactctgcccttcatgggtctgccacacaggcattctggaactgtctagctagctcttg  
45 gggaaacagctaaaaggactttggcttttctggggtttgcaggagggtaacagtgtctg  
cgcccttgtctctacttctgaatgtagtaacctcaccctctggggtagcatatgacagg  
taccacactccttttctggtgggaagcctctggcaggggagctctttctgttgcaatgtaa  
cagaggcattgcctctttcaattg

50 >scr\_sc-198205946\_1 (TOXMARKER Assignment: 116; SEQ ID NO: 116)  
gtgcacagaagtatgtgttctgggtcggaggaaagatggtaggtgtttgtcccaacacag  
tgaaaagggaacagacatgtgaagtcttcagactgtgggcctttgatttaccctcagttg  
gtctatgtgtgtaca

55 >scr\_sc-2573087\_1 (TOXMARKER Assignment: 117; SEQ ID NO: 117)  
caattgcattgcaaaattttaagggttacattgaaaacacttgaaaataagccaccaata  
aatgagatgacgataataagagccctaaataaagaggctaagaaggagttaagtgtaaa  
ggaagagggaagaaatagtttaaggcatttataagacactagaaagtctagaagagagaat  
gttagcagtagcggagtcacagctaaaaatctgcatcttgcccttttaaaaccaagagaga  
60 aagctt

5 >scr\_sc-2585074\_1 (TOXMARKER Assignment: 118; SEQ ID NO: 118)  
 agatctgctggtgtttgcctccacagtgggtgaggttgcattgacatgccgaccatgctcc  
 tatctttcacatgagtgtgtggaatgctcaggtcttagtgcttgacaagcaccttact  
 caactgaaccattgtcttagcccaatagtgaacactgaaaagttattttacccatgatc  
 agaagctttaacaatcaactagt

10 >scr\_sc-8571871\_2 (TOXMARKER Assignment: 119; SEQ ID NO: 119)  
 cctaggtctgccagtgaataagaagaccctccccgaaagtccccgagtttatgttccat  
 gcgctattcaatagccttcacatgcacatatctgcaacttcacattgatagcacagaattc  
 catcataagcatcaccatggtagccatgggtcaacaacacggaccagccatcccacctcaa  
 tagctctactgaatgggttctctgatgggttaaacgggtgatcaacatgaagctt

15 >scr\_sc-87731837\_1 (TOXMARKER Assignment: 120; SEQ ID NO: 120)  
 tgtacacaggtagtcttaggatttctgttgctgaaaccgtgggaaggggaacagttcaatg  
 agtaaaaccaagacagaagtcaacctgggttagaagctggaggcaggagaagatgcagagg  
 ctgtggaggggtgctgcttactggcttgctccccatggcttattcctgctttcttataga  
 acccaggaccaccggcccaagggttacaccatctgtggtgatctgggcccctcctccatca  
 accactaattaagaaagtgtccaagtttggctatatcttacagagatgttttctcaattg

20 >scr\_sc-87869413\_1 (TOXMARKER Assignment: 121; SEQ ID NO: 121)  
 cctaggggaatttgccattgttttagtttaagctaacactccaaaggtaatctcctatttcc  
 tcttttctcttctgtcctccatgtgggtgtcatgggcatgcagcataccagttctcaggt  
 gcctggaacactggccagtgtcttagcccagccactgtgccctgaaatccttccctgtgt  
 tcaatgctacagcacatcctccagactgcctccccacccccagcaaccgaattgagcagg  
 25 gacactaagacagtcctttggagacttccactggctgtgtgaaactttggctgctctcac  
 agcatagctcctcttagcctgtaacttagtgctgctcaggctgactgatca

30 >scr\_gb-ai233262\_2 (TOXMARKER Assignment: 122; SEQ ID NO: 122)  
 ttttttttttttttttaaggggccaagcagaagacaagctgcctttattatagttgatg  
 tcacagctctgcttgtaatagattcagccccagaaacaccccggttaaaacagcacgggt  
 gacttcaatggatagagtcttttggttaaggtgaaccagaccagggtgaccgacaatcttc  
 gggcccttgggccaggggtagcctgtagtcttacgtgaggcccagcatggcctgaagttc  
 ccgagctttatcatctggcagagagcccagggtgtgtggaagctgtcgctgtgctgctt  
 ggccaggaacgtcagtagtagtagcagtgccgcttggtgtctgggggatcctgttgct  
 35 tggcaggatcaggctgcagatgcgcaggagctctgaagccacacccacaacctgggtcagg  
 gttgtctggtgcaggaagctgaagaggtgacctatagtgaaccattcctccatgtcttc  
 cttcaggggaggggcatgtagcagggtagctagcacctggggctctgttttctcgcgg  
 actggccatcagcagacgggcaagagccccacagatgttatcacggactcgatcatgccg  
 ctcccttgccaggaggggcaaaaggaggccagtagcttagggaagtggctcctgagcagg  
 40 gcagcccccatgctctgcaagtacgcccagcccaaagatggcattgctccgcacctcggg  
 gtctgcttccccgggcattgtttaacagcacaggaaacagccgggacacaaattgggctga  
 ggcagcacctagacctgaatggattctgccagtgtccccactgcaaaggacttctctgc  
 cactgtacagctctgtttcgtcttacacagcaataatggcaacctcgtg

45 >cgrs0h0310.9\_13952-135 (TOXMARKER Assignment: 123; SEQ ID NO: 123)  
 tgatcaagggcgacacatctggagactataagaaggccctgctgctcctctgtggaggcg  
 aggatgactgaggagctgcctggagtgcctgggcccgcctgctgcccaccatcagcttc  
 cttcagcaccacgcctacttacgttcaatgcctgcctgctgccacgctgccttactcac  
 acgagtgtgtgctaatagaccaaagctgtctcgaatgaaagcagtggtctgctgttctgtc  
 50 tgacatagaccttcccacgtctctcagctagtatctctaagttgcgttttctatcctct  
 tctaaagctt

55 >scr\_gb-m13100.5\_2 (TOXMARKER Assignment: 124; SEQ ID NO: 124)  
 aagctctgggttgcttgacattgttgatcatatagggtctcgagcccccttagagctcgcc  
 agttctttctctgattccttcaacgggggtcctattctcagttcagtggtttgctgctgg  
 cattcacctctgtatttgctgtattctggctgtgtctctcaggagagatctacatccggc  
 tctgttggtctgcacttctttgcttcatccatcttgcttaattgggtggctgtatatgt  
 atgggccacatgtggggcaggctctgaatgggtgttccttctgcctctgttttaattctt  
 gctctctcttccctgccaaagggtattcttgttcccccttttaagaaggagtgaagcatt  
 60 cacattttgatcatccgtcttgagtttcatttgttctgtgcatctagggttaattcaagca  
 tttgggctaataagccatttatcaatgagtgcataccatgtatgtcttctgtgattgggt

tagctcactcaggatgatattttccagttccaaccatttgcctacgaatttcataaaactc  
 gttgtttttgatagctgagtaataattccattgtgtagatgtaccacattttctgtatcca  
 ttctctgttgaagggcatctgggttctttccagcttctggctattataaataaggctgc  
 5 aatgaacatagtgaggacacgtgtctcttttatatgttggggcatcttttgggtatatgcc  
 caagagaggtatagctggatcctcaggcagttcaatgtccaattttctgaggaacctcca  
 gactgattttccagaatgggtgtaccagtttgcaatcccaccaacaatggaggagtggtcc  
 tctttctccacatcctcgccagcatctgttgctccctgagtttttgatcatagccattct  
 cactgggtgtgaggtgaaatctcacgggtgttttgatttgcatttcccttatgactaaaga  
 10 tgttgaacattttcttaggtgtttctcagccatttggcattcctcagctgtgaattcttt  
 gtttagctctgaacccattttttaatagggttatttgtttccctgcggtctaacttctt  
 gagttctttgtatattttggatataaggcctctatctgtttaggattggtaaagatatt  
 ttcccaatctgttgggtgcccgttttgccttaaccacagtgctcctttgccttacagaagct  
 ttgcagttttatgagatcccatttgcgattcttgatcttagagcataagccattgggtgt  
 15 tttgttcaggaaatttttccagtgcccatgtgttccagatgcttccctagttttcttc  
 tattagtttgagtgtgtctggtttgatgtggaggtccttgatccacttggacttaagctt  
 tgtacagggtgataagcatggatcgatctgcattcttctacatggttgcctccagttgaa  
 ccagcaccatttgcctgaaaatgctatctttttccattggatgggttttggctcctttgtc  
 aaaaatcaagtgaccataggtgtgtgggttcatttctgggtcttcagttctattccattg  
 gtctatctgtctgtctctgtaccaatcaccatgcagtttttatcactattgctctgtaat  
 20 actgcttgagttcagggatagtgattccccctgaagtccttttattgttgaggatagctt  
 tagctatcctgggttttttgttattccagatgaatttgcaaattgttctgtct  
  
 >scr\_sc-170396977\_1 (TOXMARKER Assignment: 125; SEQ ID NO: 125)  
 tgatcacgctcagcccttggtaggacattctacagagtctcttgcctgcccctccgtctgt  
 25 gccagtggtaccacacggggcagcctccgtggaagtttctagttcacagtatgcagctca  
 gagtgaagtggtgcatcaagactccagtgctcctggaatgccagtacaaactccagg  
 cccagttcaaggacagaattacagtgctctgggattcaaaccaacagtgctgtcagtgta  
 cagccagtgatttctcctgcccattctcaagcaaccataattaccaaggacagacatgttc  
 aactgtctacgggtgtgacctctccttattcacagacaactcctccaattg  
 30  
 >scr\_sc-14059147\_2 (TOXMARKER Assignment: 126; SEQ ID NO: 126)  
 gctagcatcgtgatggccaagtgcacccctgtgcttttttcttttctaagaaagattgaa  
 aaccaacagttcttccccaacagctgcctaaattttaaggggtctgacccttacatttca  
 35 attgggggaatgaagggggcccaaccggcctaattgctgtgggagagtgagctcggatgt  
 ctgagagagcaccttgggagggactcttccctgcaatgctgtaaatacgagtaccgtttta  
 ataaagcatgtaca  
  
 >scr\_sc-87750810\_1 (TOXMARKER Assignment: 127; SEQ ID NO: 127)  
 ttttttttttttttttttggctcctgccatcttttttattgggtctgggctgtgggctgggg  
 40 gaggcaggtgggctcacatctttatgcaagcagcaaggagacgggttcacatgctcaggag  
 actccaggaaggccttgagcttgggtcgggctttgagacgcgctacataggcggagagca  
 gggggaagtctttcaagtaaccaggaacaggagctctaggttcagaagtaaatccagta  
 ggcggtagtcggcgaaggagatctggtcaccaacaatgaagcattggccacccttgttct  
 45 gggccagaagagtttcaaatggcttcagggtgcctggaagctccttcctatattggccct  
 tgtcctccttacagatatggagatagtgccatgcaatgcgcctgaacacgtcttccagtc  
 cgtcgttcaccatgtccaccagtgctgcctcttgcctgttggccttagagcccgaagg  
 agtggcccaggtgccgtaggatggcattcgattggtacagagtgagctttccatcctgga  
 acttggggatctgcccacacagacaggaagccttgaaatgtgccttgctcccaaacatcca  
 50 aggtcaccacctcctccttccaactctggccctgggtcggctagcagcatgcgcataacct  
 cacagcgcctcagtggttgggggtgcaggatggggatgaggccacagcgaagagaccacccct  
 cagagcatcctgggagagtttgggagactggaaagctgacaagtggaactaaactagcttg  
 ggagcctcgaagggagggaataatgtggtggttagaggccatgtcctaactatcttgg  
 caagccaagaccagccccaccggcacagggaaggaggaaaagtgacagacagtgtagct  
 55 gcctatggaggctaagaggtcagtcctggccccaccaaccaatttgtagtcccgcccca  
 agtctcgggtcttggccccaacgtggtcttggccacatccctccagcaccagtggtgaggg  
 ggccccaggagtgactatggccttgccttcatcttgaaaaac  
  
 >cszr\_2 02034260\_190929676 (TOXMARKER Assignment: 128; SEQ ID NO: 128)  
 gtgcaccagtagctgatgctgggagatgaatggcttagcgctgttctacttggaaacatat  
 60 cactcctgccagccgggcaactaacaattatcacccaatccaggacttaaaactgtgataga  
 ctggctgatgtttgcctttgaatagagtgctcccaaaagatgggaccactgggtcagctgcc

atggactagattctccacctgttgggggcaatctgggtcaccttgctgccaatccgacct  
 ggagccaccacagcagcagtggtcaagcactggcagaagcccatgggtggaggaaagacct  
 ctgcgactgggtgattgacctctgctgaaagccgaggctacctgtccacagacgggaac  
 agttctcttcatga

5

# **Table 9B**

>d90282 (TOXMARKER Assignment: 35; SEQ ID NO: 129)

10 aagcaaccttaaaatgactgcacctcccagatttcttttacattaactaaaaagtctta  
 tcacacaatctcataaaatztatgtaatttcatttaatttttagccacaaatcatcaaaat  
 gacgaggattttgacagctttcaaagtggtaggacactgaagactgggttttggttttac  
 caatgtgactgcacaccaaataatggaattttcaagacctggcatcaggctcctttctgt  
 caaggcacagacagcacacattgtcctggaagatggaactaagatgaaaggttactcctt  
 15 tggccatccatcctctgttggctgggtgaagtgggttttaatactggcctgggagggtaccc  
 agaagctattactgacctgcctacaaaggacagattctcacaatggccaacctattat  
 tgggaatggtaggctcctgatactacttctctggatgaactgggacttagcaaatattt  
 ggagtctaataatggaatcaagggtttcagggttgctgggtgctggattatagtaaagactacaa  
 ccactggctggctaccaagagtttagggcaatggctacaggaagaaaagggttcctgcaat  
 20 ttatggagtggacacaagaatgctgactaaaataattcgggataagggtaccatgcttgg  
 gaagattgaatttgaagggtcagcctgtggattttgtggatccaaataaacagaatttgat  
 tgcctgaggtttcaaccaaggatgtcaaagtgtacggcaaaggaaacccccacaaagtgg  
 agctgtgactgtgggattaaaaacaatgtaatccgctgctagtaaaagcgaggagctga  
 agtgcacttagttccctggaaccatgatttcaccaagatggagtgatgagggttttgat  
 25 cgcgaggagaccggggaacccagctcctgcagaaccactaatcagaatgttcagaagat  
 tttggagagtgcgcaaggagccattgtttggaatcagtagcaggaaacttaataacagg  
 attggctgctgggtgcaaaacctacaagatgtccatggccaacagagggcagaatcagcc  
 tgttttgaatatcacaaacaaacaggctttcattactgctcagaatcatttgctatgcctt  
 ggacaacaccctcctgctggctggaaccactttttgtgaatgtcaacgatcaaaacaa  
 30 tgaggggattatgcatgagagcaaaccttcttcgctgtgcagttccaccagagggtcac  
 cccggggccaatagacactgagtacctgtttgattcctttttctcactgataaagaaagg  
 aaaagctaccaccattacatcagctttaccgaagccagcactagttgcatctcgggttga  
 gggttccaaagtccttattctaggatcaggaggtctgtccattgggtcaggctgggaatt  
 tgattactcaggatctcaagctgtaaaagccatgaaggaagaaaatgtcaaaactgttct  
 35 gatgaacccaaacattgcatcagtcagaccaatgaggtgggttaaagcaagcggatac  
 tgtctactttcttcccatccccctcagtttgcacagagggtcatcaaggcagaacagcc  
 agatgggttaattctgggcatgggtggccagacagctctgaactgtggagttagaactatt  
 caagagaggtgtgctcaaggaatatgggtgtgaaagtcctgggaacttcagttgagtcct  
 tatggctacggaagacaggcagctgtttcagataaaactaaatgagatcaatgaaaagat  
 40 tgctccaagttttgcagtgggaatcgattgaggatgcactgaaggcagcagacaccattgg  
 ctaccagtgatgatccgttcgcctatgcactgggtgggttaggctcaggcatctgtcc  
 caacagagagactttgatggacctcagcacaaaggcctttgctatgaccaaccaaatct  
 ggtggaagactcagtgacaggttggaaagaaatagaatatgaagtgggttcgagatgctga  
 tgactcagttgttgggtcctgcccagacactctccaatgccgagtttcagatgttgag  
 45 acgtacttcaatcaatgttgttcgccacttgggcattgtgggtgaatgcaacattcagtt  
 tgcccttcacctacctcaatggaatactgcacattgaagtgaatgccaagatgtcccc  
 gaactctgctctggcctccaaaacgactggctaccattggcattcattgctgcaaaagat  
 tgccctaggaatcccacttccaggaattaagaacgtcgatccgggaagacatcagcctg  
 50 ttttgaacctagcctggattacatgggtcaccaagattccccgctgggatcttgaccgttt  
 tcatggaacatctagccgaattggtagctctatgaaaagtgtaggagaggtcatggctat  
 tggctgtacctttgaggagaggtttccagaaagccttacggatgtgccaccatctataga  
 ggggtttcactccccgtctcccaatgaacaaagaatggccatcgaatttagatcttagaaa  
 agagttgtctgaaccaagcagcagcgtatctatgccattgccaaggccattgatgacaa  
 55 catgtcccttgatgagattgagaagctcacatacattgacaagtgggtttttgtataagat  
 gcgtgatatttttaacatggaaaagacactgaaaggcctcaacagtgaagtcctatgacaga  
 agaaaccctgaaaagggcaaaggagattgggttctcagataagcagatttcaaaatgcct  
 tgggctcactgaggccagacaaggagctgaggttaaagaaaaacatccacccttgggt  
 60 taaacagattgatacactggctgcagaatacccatcagtaacaaactatctctatgttac  
 ctacaatgggtcaggagcatgatgtcaattttgatgaccatggaatgatgggtgctaggctg  
 tgggtccatcacattggcagcagtggtgaatttgattgggtgtgctgtctctagtatccg

5 cacactgcggtcaacttggcaagaagacggtgggtggaattgcaatcctgagactgtgag  
 cacagactttgatgagtgacaaactgtactttgaagagttgtccttggagagaatcct  
 agacatctaccatcaggagcatgtggtggctgcatcatatcagttggaggccagattcc  
 aaacaacctggcagttcctctatacaagaatggtgtcaagatcatgggcacaagccccct  
 gcagatcgacagggtgaggatcgctccatcttctcagctgtcttggatgagctgaaggt  
 ggctcaggcaccttggaaagctgttaatactttgaatgaagcactggaatttgcaaagtc  
 tgtggactaccctgcttgttgaggccttccatgttttgagtgggtctgctatgaatgt  
 ggtattctctgaggatgagatgaaaaaattcctagaagaggcgactagagtttctcaggc  
 10 cacgccagtggtgctgacaaaatttgttgaagggggcccgagaagtagaaatggacgtgt  
 tggcaaagatggaaggggttatctctcatgccatctctgaacatgttgaagatgcaggtgt  
 ccactcgagaaatgccactctgatgctgccacacaaaccatcagccaagggggccattga  
 aaagtggaaggtgctaccggaagattgcaaaggcttttgccatctctgggtccattcaa  
 cgtccaatttcttgtcaaaggaaatgatgtcttgggtgaatgagtgtaacttgagagcttc  
 15 tcgatccttccccctctgtttccaagactcttggggttgacttcattgatgtggccaccaa  
 ggtgttgattggagagaatgttgatgagaaacatcttccaacattggaccatcccataat  
 tcctgttgactatgttgcaattaaggctcccatgttttccctggccccgggtgagggatgc  
 tgacccattctgagatgtgagatggcttccactggagaggtggcttgcttgggtgaagg  
 tattcatacagccttccataaaggcaatgctttccacaggatttaagataccccagaaagg  
 catcctgataggcatccagcaatcattccggccaagattccttgggtggtggaacaatt  
 20 acacaatgaaggtttcaagctgtttgccacggaagccacatcagactgggtcaacgcca  
 caatgtccctgccaaccagtggtgcatggcgtctcaagaaggacagaatcccagcctctc  
 ttccatcagaaaattgattagagatggcagcattgacctagtgattaaccttcccaacaa  
 caacactaaatttgtccatgataattatgtgattcggaggacagctgttgatagtggaat  
 cctctcctcactaattttcaggtgaccaaactttttgctgaagctgtgcagaaatctcg  
 25 caagtgagactccaagagtctttccactacaggcagtagctggtgaaaagcagcata  
 gagatgcagacacccagccccattattaaatcaacctgagccacatgttatataaaggga  
 actgattcacaaactttctcagagatgaatattgataactaaacttcatttcagtttactt  
 tgttatgccttaataattctgtgtcttttgcaattaaattgtcagtcacttcttcaaaacc  
 ttacagtccttccctaaggttactcttcatgagattcatccatttactaataactgtat  
 30 tgggtgactaggcttgccatgtgtcttatgtgtagctttttactttttatggtgtgatta  
 atggtgatcaaggtaggaaaagttgtgttctattttcttgaactccttctatactttaag  
 atactctattttttaaaactatctgcaaaactcaggacactttaacaggggcagaatactc  
 taaaaacttgataaaattaaatatagatttaatttatgaaccttccatcatgtgtttgtg  
 35 tattgttctttttggatcctcattctcaccatttggctaattccaggaatattgttatc  
 cttcccattatattgaagttgagaaatgtgacagagcatttagagtatgaattc

>xm\_114110 (TOXMARKER Assignment: 71; SEQ ID NO: 154)  
 40 gtgcccgggaagtggctccagggagaagaggcctcttccctcaccgcgtgtgggagctgcgccccgaaagcct  
 gccccggcacgtcggtctctctgaccgccaagaccagagagccgttggcgccctccgccccggcctgccc  
 gtccgtttattttaagaagctttgtgcgcctgctgtggggatttctgatccaggctgcgaagaatttcgaag  
 tctggaaaatagcaactgtgtttgtttctaaaggatcttctcctgaccagcatcgctcatcacaatgaaga  
 accaagacaaaaagaacggggctgccaacaatccaatccaaaaagcagcccaggacaaccggaagcaggac  
 ccgagggagcccaggagcgggccagccaggcggtcctgcagtagaagcagaaggtcccggcagcagccagg  
 45 ctctcggaagccggagggggtcaagccagaacggctcagctcggggcccttcgtgatgtctctgaggagc  
 tgagccgcaactggaagacatactgagcacatactgtgtggacaataaccaggggggccccggcgaggatg  
 gggcacaggggtgagccggctgaaccggaagatgcagagaaagtcccggacctatgtggcaaggaatggggagc  
 ctgaaccaactccagtagtcaatggagagaaggaacctccaaggggggatccaaacacagaagagatccggc  
 agagtgcagaggtcggagaccgagaccatcgaaggccacaggagaagaaaaagccaaggggttgggggaagg  
 50 agatcacgttgctgatgcagacattgaatactctgagtacccagaggagaagctggctgctctgtgcaaga  
 agtatgctgaactgctggaggagcaccggaattcacagaagcagatgaagctcctacagaaaaagcagagcc  
 agctggtgcaagagaaggaccacctgcgcggtgagcacagcaaggccgtcctggccccgagcaagcttgaga  
 gcctatgccgtgagctgcagcggcacaaccgctcctcaaggaagaaggtgtgcagcggggccgggaggagg  
 aggagaagcgcaaggaggtgacctcgacttccaggtgacactgaatgacattcagctgcagatggaacagc  
 55 acaatgagcgcaactccaagctgcgccaaagagaacatggagctggctgagaggctcaagaagctgattgagc  
 agtatgagctgcgcgaggagcatatcgacaaagtcttcaaacacaaggacctacaacagcagctggtggatg  
 ccaagctccagcaggcccaggagatgctaaaggaggcagaagagcggcaccagcgggagaaggattttctcc  
 tgaaagaggcagtagagtcccagaggatgtgtgagctgatgaagcagcaagagacccacctgaagcaacagc  
 ttgccctatacacagagaagtttgaggagttccagaacacactttccaaaagcagcgaggtattcaccacat  
 60 tcaagcaggagatggaaaagatgactaagaagatcaagaagctggagaaagaaaccacctgtaccggtccc  
 ggtgggagagcagcaacaaggccctgcttgagatggctgaggagaaaacagtcggggataaagaactggagg  
 gcctgcaggtaaaaatccaacggctggagaagctgtgccgggcactgcagacagagcgcaatgacctgaaca



cctgcgcaagcagcagaggcgcgagctgctcttcaacgaggtggtaatcatgagggacta  
 ccagcacgagaatgtggtggagatgtacaacagctacctgggtgggggacgagctctgggt  
 ggtcatggagttcctggaaggaggcgccctcaccgacatcgtcaccacacaccaggatgaa  
 cgaggagcagatcgcgccgctgtgccttgagtgctgcagggcctgtcggtgctccacgc  
 5 ccagggcgctcatccaccgggacatcaagagcgactcgatcctgctgacccatgatggcag  
 ggtgaagctgtcagactttgggttctgcgcccagggtgagcaaggaaagtgccccgaaggaa  
 gtgcgtgggtcggcacgcccactggatggccccagagctcatctccgccttccctacgg  
 gccagaggtagacatctggtcgctggggataatggtgattgagatggtggacggagagcc  
 10 cccctacttcaacgagccacccctcaaagccatgaagatgattcgggacaacctgccacc  
 ccgactgaagaacctgcacaagggtgtcgccatccctgaagggttctctggaccgctgct  
 ggtgcgagacctgcccagcgggcccagcgagcagctgctgaagcaccatctcctggc  
 caaggcagggccgctgcccagcatcgtgcccctcatgcgccagaaaccgcaccagatgagg  
 cccagcgcccttcccctcaaccaaagagcccccggtcacccccgccccactgaggcca  
 gtagggggccaggcctcccactcctcccagcccgggagatgctccgcgtggcaccacct  
 15 ccttgctggggtagatgagaccctactactgaactccagttttgatctcgtgactttta  
 gaaaaacacagggactcgtgggagcaagcgaggtcccaggacccccaccctctgggaca  
 ggccctccccatgttcttctgtctccaggaaggggcagcgccctcccatcactggaagt  
 ctgcagtgggggtcgctgggggtggagagaacactaagaggtgaacatgtatgagtgtgt  
 gcacgcgtgtgagtgtgcatgtgtgtgtgtgcaaagggtccagccaccccgctcctccagcc  
 20 tgcaaggggtgtctggcgcttgccgtgacaccagccccctctccccctgagccattgtg  
 ggggtcgatcatgaatgtccgaagagtggccttttcccgtagccctgcgccccctttctg  
 tggctggatggggagacaggtcagggccccccaccctctccagccctgcagcaaatgac  
 tactgcacctggacagcctcctcttttctagaagtctattttatattgtcattttataaca  
 ctctagcccctgcccattattgggggacagatggtccctgtcctgcgggtggccctggca  
 25 gaaccactgcctgaagaaccaggttccctgcccgggtcagcgcagccccagccccacc  
 ctgcctcgagttagttttacaattaaaacattgtcttgtttgtgtctgtgtgcgatgtg  
 tggggggcagggggccctgcccggctgtcttgggtgggaatttgcagggagaggggtctgg  
 atctggggagcaaacacgattccagccaaggcagggcaagggtgggggtggggagtgggga  
 gttcaggtcatagcagccagtaagctccccagcctgcccactcccagaatggggcagga  
 30 ttgtccccaccctggaagcagccagtttgccacagtccatgtgcagactgatcccagtt  
 tgccaaatctgcaatttctggaaccttttaaaggctgtcttgagcgcgtttgggtgagta  
 ggagctaaccacagtttagtaaatgaaggccatttgcaaatgggtcagtgggcagatgg  
 gcttttggggattgactgaggctgactggcctggagctgctggcttcggagagacacct  
 gtgaagtgtgtccttccacgcaggagcccagagccgagcccacgctggggggaatctgac  
 35 tggcatggaggtggccatgccaccatcgctgctgcagctgcatcctggcactttgcccct  
 caggccctgtgggctccactttctgcatcctcccagccccagggaggcagtgaggatg  
 gggagagagccaggagtgagcctccgtccccaaagccagccaggcgtcatcagcaccaga  
 gacctcagcctggtcctctcgggaaagtgagtggccagggcagagattccaggttagtcca  
 cgcctcccacccttcacaggtcctgaccccaagaatcagagcactgtgtgtgtggcaggg  
 40 cctatgccaaagtgcaaacacagcctagatggatcatcacagagtgaacccagcggtgca  
 agcagctgtgctctctgcgatgtattggaggcttaggtgaggtggatgcctttctggaaa  
 aaaaaaaaaatgctaacattggcaaaaagaagaatagaaaaacaagacaaaaataactgtct  
 cctcactgcacacacactccagaataaaataaaaagggttcaggcttgaatgcactttcaaa  
 tgagattttttttttttttttttttgagacggagtatcgctctgtcgcccaggctggagtgcg  
 45 tggcacggctcttggctcgctgcaacttctgcctcctgggttcaagcgatttctcctgcctc  
 agcctcccagggtggctgggattgcaggcgcttgccaccacgcccgggctagtttttgtat  
 tttgggtagagactgggtttcgccatcttggccagactggcttgaactcctgacctcgt  
 gatccgcccacctcagcctcccaagggtgctgggattgcaggcatgagccaccatgcctgg  
 cctcaaatgaggtttaccagactttgaaggagcaggtaatccttctaccttgtgaacaa  
 50 gtcgttccagaaagatagcagctcaggaggcctctgtgaccatgggtccagaccagata  
 aggacggcgaagaacagagcatctcagaaacgcaaggctcacagccagggtgcccggcga  
 cccacgggactgagaacagctagctctaggagctccactctcctgctgaagaaaccac  
 gggctcagagacggggagctccctcgccagccacatctgtgacccacggtaactctgc  
 tggtttttggctccttcagtcactcactgcaggtttgttttgttttgttttgtttt  
 55 gagttttttgttttttgtttttgttttttagagacgggatgttgctatgttgccctaggc  
 tggtttcgaactcctggcctcaggggatcctcctccttggccccgcaagggtgctgggat  
 tgcaggcgtgagccgctgcccggcccgtcactgcagtttgaaggcatggctttgggt  
 ggcgtgggggtgaaagctgcccagggccccgttccctcccacgtggctgcctcctgccaga  
 gccagttaggaaaacagaccccaactagagttgtttcaaatggcagggatttgggtaccgg  
 60 tgggtggatcatgacaaagctctgagaaggctggaggagccacagagtgccaaagtgccca  
 gcaatcattagaggaaggaggctgctgccacctgtgtggctagaggaacagaggggccaa

5 tggcattccccaaaccccactctcgccctctgtctgcccagagcagaatggcttcttccag  
 cttcccaccctggactcccaccaggagcctcctcctggcagacccttccctgacccacc  
 ggcccggggggtctacagatccatgttttcaggcgtccgcctggagcggaacaggggagtg  
 cttaggacaaggggtggtgccagaggatccactctgccacatttagttgaccagctgagg  
 10 cactccacgggaatgaatgactctcgacaggtgccggaggtgaggaggggcccggaggcc  
 caggaggggcacagggatggattcgctccgcctgggggctggaggtgtgtttacagagccc  
 caaaataaacaatgcaaccagggtcagaccagcggttctcacacagtgtggttcccagacc  
 agcagcatcactgggagcttactggacacgcaaataaatccctgtgtgccacccagctg  
 catcagatgctgggtggggcccagtgatctgtatttaacacaccctccgggggatgccggt  
 15 gccactcacgtttgagaacccctgcatccacgactgccctcccgtgtaaaaggcccac  
 ctctgtgggactccaagtcacagcaccctagggtccttccgtctttttccttccctcctg  
 ggacacctgcctctcccatgtcgtattagagaattccttatgctcccaagtgggcacggg  
 gagaggaaggcactcctctttaaggaccgacccagagggttttgccattgcttcaactggcc  
 agagcttagtcacgcagcctcaccagaggcaaggaggttggaatatgtagtgtttgtg  
 20 tgtgtctaacacaaattctattaccatgcagtcaggattctccactcttgctctttcatt  
 agatttgcctgggcttaccctggacttctgatttagtgacagaacagagaacccagagg  
 cagaccagatgtgtacaagggttcatatacaatcaggagatttaataatcatgctagg  
 ggccgggtgcagtggtcacgcctgtaatcccaagcactttggggagccgaggcaggcgg  
 atcacttgaggtcaggagtttgagaccagcctgggcaacaaagtgcagccctgtctctac  
 25 taaaaataacaaaaattagccgggctggtggtgggtgcctgtaatcccagctcctcggg  
 tggctgaggcatgagaatcacttgaacccaggaggcagaggtttcagtgagctgagatca  
 catcactgcactccagcctgggtgacagagtgcagattccgtc  
  
 >m14328.1 (TOXMARKER Assignment: 37; SEQ ID NO: 168)  
 30 acggagatctcgccgctttacgttcacctcgggtgtctgcagcaccctccgcttccctctc  
 ctaggcgacgagaccagtggttagaagttcaccatgtctattctcaagatccatgccag  
 ggagatctttgactctcgccggaatcccactgttgaggttgatctcttcacctcaaaagg  
 tctcttcagagctgctgtgccagtggtgcttcaactgggtatctatgaggccctagagct  
 35 cgggacaatgataagactcgctatatggggaagggtgtctcaaaggctgttgagcacat  
 caataaaactattgcgcctgccctgggttagcaagaaactgaacgtcacagaacaagagaa  
 gattgacaaactgatgatcgagatggatggaacagaaaaataaatctaagtttgggtgcgaa  
 cgccattctgggggtgtcccttgccgtctgcaaagctgggtgccgttgagaaggggtccc  
 cctgtaccgccacatcgctgacttggtggcaactctgaagtcacctgccagtcgggc  
 40 gttcaatgctcatcaatggcggttctcatgctggcaacaagctggccatgcaggagttcat  
 gatcctcccagtcgggtgcagcaacttcagggaagccatgcgcattggagcagaggttta  
 ccacaacctgaagaatgtcatcaaggagaaatatgggaaagatgccaccaatgtggggga  
 tgaaggcgggtttgctcccaacatcctggagaataaagaaggcctggagctgtgaagac  
 tgctattgggaaagctggctacactgataaggtgggtcatcgccatggacgtagcggcctc  
 45 cgagttcttcaggctctgggaagtatgacctggacttcaagtctcccgatgacccagcag  
 gtacatctcgctgaccagctggctgacctgtacaagtccttcatcaaggactaccagt  
 ggtgtctatcgaagatcccttgaccaggatgactggggagcttggcagaagttcacagc  
 cagtgcaggaatccaggtagtgggggatgatctcacagtgaaccaaccaaaggatcgc  
 caaggcgtgaaacgagaagtcctgcaactgcctcctgctcaaagtcaaccagattggctc  
 50 cgtgaccgagctctcttcaggcgtgcaagctggcccaggccaatggttggggcgtcatgggt  
 gtctcatcggttgggggagactgaagataccttcatcgctgacctggttgtggggctgtg  
 cactgggcagatcaagactggtgcccttgccgatctgagcgttggccaagtacaacca  
 gctcctcagaattgaagaggagctgggcagcaaggctaagtttgccggcaggaacttcag  
 55 aaaccccttgccaagtgaagctgtgggcaggcaagccttcggtcacctgttggtacaca  
 gacccctcccctcggtgcagctcaggcagctcgaggcccccgaccaacacttgcaggggt  
 cctgctagttagcgccccaccgcccgtggagttcgtaccgcttccctagaacttctacag  
 aagccaagctccctggagccctgttggcagctctagcttttgagtcgtgtaatgggccc  
 60 aagtcattgtttttctgcctcactttccaccaagtgtctagagtcagtgtgagcctcgtg  
 tcactcgggggtggccacaggctagatccccgggtggttttgcgtcaaaaataaaaagcc  
 tcagtgacctatgag  
  
 >xm\_035635 (TOXMARKER Assignment: 68; SEQ ID NO: 169)  
 gggggacgggtgaaggttgccctcccgcccgtccgggctctgatcctccgtctccccgtccccggcgccggc  
 ccatggcctggcggaggcccgaacatggacctccgcaccgcccgtgtacaacgcccgcctgatggcaagct  
 gcagctgctccagaagctgctcagcggccggagccgggaggaaactggacgagctgacgggcgagggtggccgg  
 65 cgggggaacgcccgtactcatcgcccccgtacggccacctggacgtggtggagtacctggtggaccgggtg  
 cggcgcgagcgtggaggccggtggctcgggtgcacttcgatggcgagaccatcgagggcgccgcccgtgtg



5 ggccgcctccgcagccggccacctggacgtggtgcggagcctgctgcgcgcggggcctcggatgaaccgcac  
 cacgcgcaccaactccacgcctctccgcgcgcgcctgcttcgacggccacctggaggtggtgcgctacctggt  
 cggcgagcaccaggccgacctggaggtggccaaccggcagccacacgtgacctcatgatctcgtgctacaa  
 10 gggccaccgtgagatcgcccgctacctgctggagcagggcgccaggtgaaccggcgagcgccaagggcaa  
 cacggccctgcatgactgcgcgagtcggcgagcctggagatcctgcagctgctgctggggtgcaagggccg  
 catggaacgtgacgggtacggcatgaccccgctgctcgcggccagcgtgacggggccacaccaacatcgtgga  
 gtacctcatccaggagcagcccgccaggagcaggtcgacgggggagaggctcagcctgggctgccccaaaga  
 agacccctccaccagccaggggtgtgcgcagcctcagggggctccgtgctgcagctcctccccagaggaacc  
 actgaacgggggaatcttacgaaagctgctgtcccaccagccgggaagctgccgtggaagccttggaattgct  
 15 gggagctacgtatgtggataagaaacgagatctgcttggggcccttaaacactggaggcgggcatggagct  
 gcgtcaccaggggggagtagtacctgccccaaaccggagccccacagctggtcctggcctatgactattccag  
 ggaggtcaacaccaccgaggagctggaggcgctgatcaccgacccggatgagatgcgcatgcaggccctgtt  
 gatccgggagcgcacacctcggtccctcgaccccgacacttctattacatccgttacaggggtgccgtgta  
 cgccgactcgggcaatttcgagcgtgcatccgcttggtgaagtagccctggacatgcaacagagcaacct  
 20 ggagcctctgagccccatgaccggcagcagcttctctccttcgcggaactcttctctacgtgcttcaggga  
 ccgggacgccaagggcagcctgggaccccgatcggtttgcagacctcatgggggttctcaccaaaggggt  
 ccgggaagtggaaacgggcccctgcagctgcccaggagcccgagactcagcccagttcaccaaggcgctggc  
 catcatcctccacctgctctacctgctggagaaagtggagtgacccccagccaggagcacctgaagacca  
 gaccgtctaccgctgctcaagtgcgcgcccaggggcaagaacggcttcacccctctgcacatggctgtgga  
 25 caaggacaccacaaacgtgggcccgtatcccgtgggagattcccctccctgcagctgggtcaaagtgtgct  
 cgactgcggggccgacccggacagcagggattttgacaacaacaccccgctacacatagcagcccagaacaa  
 ctgcccggccatcatgaatgccctgatcgaagcagggggccacatggacgccaccaatgccttcaagaagac  
 ggcctacgagctgctggacgagaagctgctggccaggggtaccatgcagcccttcaactacgtgacctgca  
 gtgccttgccggcccgggcccctggataagaacaagatcccttacaagggttcacccggaagatctggaggc  
 30 gttcatcgaactgcactgacctgcccagaagcctgcacccctcacctctccctctcctgctgagatggggg  
 aaatccggctgcggcatagcagatgctcgttcttgcctccttcaggcaccaatcaggagaagggttctgcct  
 cccatccctctacctgcagacagggctcgagggtgttagcgagcctttgggtgctagaagcctgcggggtcat  
 gtgctaagaggacagtctttctccgggagcccgctcactcattctgagttaggaaaagacacaagaccttcc  
 ccacatcctgtctgcctgggttagggaggcctttgcctgttacctagaggcggagggactgaagccattgc  
 35 gttccttccctgctagaaacacaggaagaagttgaggactgtctgccttccctcgctccctttacctggccag  
 ataactccagccgctgaatacagtggttaggactgggggctcctgagatgagagtttgagattcagggaatga  
 gaccacctctcatcttccagcatgatcgcgccctgctcccggtgccaccgtagtccttggcagacaggcag  
 ggctctgcccagggcagcctgccacttgcatagctttcggttggttggtgttctgtttatataaagtgg  
 40 gcaggttgcaagcgttgacagaaattctgagattttactgccttttttttttttttttaagaaagtgtgtg  
 ttggtactcataagtgaatttcaagcagtgaggattttgtggtgcctgagatggccgagggcacaggagtg  
 agctgtatgtgtgaggaatttggtgagcgagataaaagtccacgggtgtcaacccctaaaacatgggtgaccg  
 tacatttttatacatctccactctacggccttttacaggctttccgattttacaggcctttccaagtttcca  
 ttctccttagagagagaactgtgcttccaaacagaaatcaggagtgaccacaaagcctgaaaacactttgcc  
 45 accagcaaaagaactggcacaattgggttggtgtgcatatgcatagtgcccgagttaaaactgcaggccac  
 tctgccttgacgtggcctctgatttctattgtgggtgcatccacaggtggcccagctgttctttcagctgct  
 ccaaggattgagaccgaagtcatcatgaaaaaggcccaagtacagctctaatgcgataaatccactagctaa  
 gacgtcgagtgccaagaccagccttccagccgaggtttggacaaagtctcaggttcccgtgactcagggttaa  
 ggtgctggggctgcccagaggacctgcccagcaagattttgtcaagagcgagactccatcagcccaggcag  
 acgggagcaggttcttggccagcgtagacagcagcaaacagcagcaggggaagccattctcactgcacccctc  
 50 ctgcagtacccaagtcaggcccttaggaggagcagtgacgggggtgtccagaaatatcctgtccctggat  
 ggaaactaggtctcgtttggatttttttttttttttggcgtgttaggaaattatttatttataagaca  
 ggttttaactcagccgaggtgggaaatggtgtccctgtccctccaaagcacagagcacagaaatgaggccg  
 ttacatggcgagctctccgtgctggtgtttaagtcattaaaaagataactcaaagag

50 >bc020744 (TOXMARKER Assignment: 55; SEQ ID NO: 170)  
 acaggatctgcttagtgaaagaagtggcaagcaatggatcccaaatatcagcgtgtagag  
 ctaaatgatgggtcacttcatgcccgtattgggatttggcacctatgcacctccagaggtt  
 55 ccgaggaacagagctgtagaggtcaccaaattagcaatagaagctggcttccgcatatt  
 gattctgcttatttatacaataatgaggagcaggttgactggccatccgaagcaagatt  
 gcagatggcagtgatgaagagagaagacatatctacacttcaaagctttgggtgactttc  
 tttcaaccacagatggtccaaccagccttggaagctcactgaaaaaacttcaactggac  
 tatgttgacctctatcttcttcttcttcccaatggctctcaagccaggtgagacgccacta  
 ccaaaagatgaaaatggaaaagtaatatcgacacagtggtatctctctgccacatgggag  
 60 gtcattggagaagtgtaggatgcaggattggccaagtccatcggggtgtcaaacttcaac  
 tacaggcagctggagatgatcctcaacaagccaggactcaagtacaagcctgtctgcaac  
 caggtagaatgtcatccttacctcaaccagagcaaatgctggatttctgcaagtcaaaa

5 gacattgttctggttgccacagtgctctggaacccaacgacataaaactatgggtggac  
 ccaaactccccagttcttttgaggagcccagttctttgtgccttagcaaagaaacacaaa  
 cgaaccccagccctgattgccctgcgctaccagctgcagcgtggggttgtggctcctggcc  
 aagagctacaatgagcagcggatcagagagaacatccagggttttgaattccagttgaca  
 tcagaggatatgaaagtcttagatgggtctaaacagaaattatcgatatgttgtcattggat  
 10 tttcttatggaccatcctgattatccattttcagatgaatattagcatagaggggtgtgc  
 acgacatctagcagaaggccctgtgtgtggatgggtgatgcagaggatgtctctatgctgg  
 tgactggacacacggcctctgggttaaatccctccctcctgcttggaacttcagctagc  
 tagatatatccatggtccagaaagcaaacataataaatttttatcttgaactaaaaaaaa  
 aaaaaaaaaaaaaaaaaa

>x52425.1 (TOXMARKER Assignment: 43; SEQ ID NO: 133)  
 15 ggcgaatggagcaggggcgcgcagataaattaaagatttacacacagctggaagaaatcat  
 agagaagccggcggtggtggtcatgcctataatcccagcacttttgagggtgagggcgg  
 gcagatcacttgagatcaggagttcgagaccagcctgggtgccttggaatctcccaatggg  
 gtggctttgctctgggctcctgttccctgtgagctgcctgggtcctgctgcaggtggcaag  
 ctctgggaacatgaaggctcttgaggagcccacctgcgtctccgactacatgagcatctc  
 tacttgcgagtggagatgaatgggtccaccaattgcagcaccgagctccgctgttgta  
 20 ccagctgggttttctgctctccgaagcccacacgtgtatccctgagaacaacggaggcgc  
 ggggtgcgtgtgccacctgctcatggatgacgtgggtcagtgcggaataactatacactgga  
 cctgtgggctgggcagcagctgctgtggaagggctccttcaagcccagcgagcatgtgaa  
 acccagggccccaggaaacctgacagttcacaccaatgtctccgacactctgctgctgac  
 ctggagcaacccgtatccccctgacaattacctgtataatcatctcacctatgcagtcaa  
 catttgagtgaaaaacgacccggcagatttcagaatctataacgtgacctacctagaacc  
 25 ctccctccgcatcgccagccagcaccctgaagctctgggatttccctacagggcacgggtgag  
 ggcctgggctcagtgctataaacaccacctggagtgagtgaggccccagcaccaagggca  
 caactcctacagggagcccttcgagcagcacctcctgctgggcgtcagcgtttcctgcat  
 tgtcatcctggcgtctgcctgttgtgctatgtcagcatcaccaagattaagaaagaatg  
 gtgggatcagattcccaacccagcccgcagccgctcgtggctataataatccaggatgc  
 30 tcaggggtcacagtgaggagaagcggtcccagggccaggaaccagccaagtgccacactg  
 gaagaattgtcttaccaagctcttgccctgttttctggagcacaacatgaaaagggatga  
 agatcctcacaaggtgccaaagagatgcctttccagggctctggaaaatcagcatggtg  
 ccagtgagatcagcaagacagtcctctggccagagagcatcagcgtggtgcgatgtgt  
 ggagttgtttgaggccccgggtggagtgtagggaggaggagggtagaggaaagaaaagg  
 35 ggccttctgtgcatcgctgagagcagcaggatgacttccaggagggaagggagggtcat  
 tgtggcccggttaacagagagcctgttccctggacctgctcggagaggagaatgggggctt  
 ttgccagcaggacatgggggagtcagccttctccaccttcgggaagtacagtgctca  
 catgccctgggatgagttcccaagtgaggggcccaaggaggcacctccctggggcaagga  
 gcagcctctccacctggagccaagtctcctgccagcccagccagagtcagacaacct  
 40 gacttgcacagagacgcccctcgatcgagggcaacctgcttaccgcagcttcagcaa  
 ctccctgagccagtcaccgtgtcccagagagctgggtccagacccactgctggccagaca  
 cctggaggaagtagaaccgcagatgccctgtgtccccagctctctgagccaacctgtg  
 gcccaacctgagccagaaacctgggagcagatcctccgccgaaatgtcctccagcatgg  
 ggagctgcagcccccgctctcgccccaccagtggtatcaggagtttgtacatgcggtg  
 45 ggagcagggtggccaccaggccagtgcggtgggtgggcttgggtccccaggagaggctgg  
 ttacaaggccttctcaagcctgcttgccagcagtgctgtgtccccagagaaatgtgggtt  
 tggggctagcagtggggaagaggggtataagcctttccaagacctcattcctggctgcc  
 tggggaccctgccccagtcctgtcccctgttcacctttggactggacagggagccacc  
 tcgcagtcgcgagagctcacatctcccaagcagctccccagagcacctgggtctggagcc  
 50 gggggaaaaggtagaggacatgccaagccccacttccccaggagcaggccacagaccc  
 ccttgtggacagcctgggcagtggtgttactcagcccttacctgccacctgtgcgg  
 ccacctgaaacagtgatggtggcaggaggatggtggccagacccctgtcatggcagctcc  
 ttgctgtggctgctgtgtggagacaggtcctcgccccctacaacccccctgagggcccc  
 agaccctctccaggtggggttccactggaggccagtcgtgtgtccggcctccctggcacc  
 55 ctcgggcatctcagagaagagtaaatcctcatcatccttccatcctgccccggcaatgc  
 tcagagctcaagccagacccccaaaatcgtaactttgtctccgtgggacccacatacat  
 gagggctctctaggtgcatgtcctctgttgtgtgagtcgtgcagatgaggactagggtta  
 tccatgcctgggaaatgccacctcctggaaggcagccaggctggcagatttccaaaagac  
 ttgaagaacatggtatgaagggtgattggccccactgacgttggcctaactgggctgc  
 60 agagactggaccccgccagcattgggctgggctcgccacatcccatgagagtagagggc  
 actgggtcgccgtgccccacggcaggccccctgcaggaaaactgaggcccttgggcacctc

gacttgtgaacgagttgttggctgctccctccacagcttctgcagcagactgtccctgtt  
gtaactgcccgaagcatgttttgcaccagatcatggcccacgtggaggcccacctgcc  
tctgtctcactgaactagaagccgagcctagaaaactaacacagccatcaagggaatgact  
5 tggcgcccttgggaaatcgatgagaaattgaacttcagggagggtggtcattgcctaga  
ggtgctcattcatttaacagagcttccttaggttgatgctggaggcagaatcccggctgt  
caaggggtgttcagttaaggggagcaacagaggacatgaaaaattgctatgactaaagca  
gggacaatttctgcccacacccatgccagctgtatggctgggggctcctcgatgca  
10 tggaaacccccagaataaataatgctcagccaccctgtgggcccgggcaatccagacagcagg  
cataaggcaccagttaccctgcatgttggcccagacctcaggtgctaggggaaggcgggaa  
ccttgggttgagtaatgctcgtctgtgtgttttagtttcatcacctgttatctgtgtttg  
ctgaggagagtggaaacagaaggggtggagttttgtataaataaagtttctttgtctc

>j02906 (TOXMARKER Assignment: 48; SEQ ID NO: 140)  
gcaggctcagcgatcccagccagtgctcctgcagctcagcagctgccttcaccatgga  
15 cagcataagcagagccatcttactcctgctcctggctctcgtctgtctgctcctgacct  
aagctcaagagataaagggaaagctgcctccgggaccagaccctctcaatcctgggaaa  
cctgctgctgctttgctcccaagacatgctgacttctctcactaagctgagcaaggagta  
tggctccatgtacacagtgcacctgggacccaggcgggtggtggtcctcagcgggtacca  
20 agctgtgaaggaggccctgggtggaccaggagaggagtttagtggccgcggtgactacc  
tgcctttttcaactttaccaagggcaatggcatcgcttctccagtggggatcgatggaa  
ggtcctgagacagttctctatccagattctacggaatttcgggatggggaagagaagcat  
tgaggagcgaatcctagaggagggcagcttcctgctggcggacgtgcggaaaactgaagg  
cgagccctttgacccacgcttctgctgagtcgctcagtggtccaacattatctgttccgt  
gctcttcggcagccgcttcgactatgatgatgagcgtctgctcaccattatccgccttat  
25 caatgacaacttccaaatcatgagcagccctggggcgagttgtacgacatcctagacc  
cagattcccagacctcctggactgggtgcctgggcccagccaacgcacatctccagaact  
caagtgcctgagagacctcatcgcccacagcgtccacgaccaccaggcctcgtctccccg  
ggacttcatccagtgcttcctcaccaagatggcagaggagaaggaggaccactgagcca  
cttcacatggataccctgctgatgaccacacataacctgctctttggcggcaccaagac  
30 ggtgagcaccacgctgcaccacgccttcctggcactcatgaagtacccaaaagtccaagc  
ccgctgagcaggaggagatcgacctcgtgggtgggacgcgcgcggctgcggcgctgaagga  
ccgcgcgccatgccttacacagacgcggtgatccacgaggtgcagcgtttgcagacat  
catccccatgaacttgccgcaccgcgtcactagggacacggcctttcgcggttcctgat  
acccaagggcaccgatgtcatcacctccttaacaccgtccactacgacccagccagtt  
35 cctgacgccccaggagttcaaccccagacattttttggatgccaatcagtccttcaagaa  
gagtcagccttcatgccttctcagctgggcgcgctctgtgctgggagagctgctggc  
gcgcatggagctcttctgtacctcacgcacatcctgcagagcttttcgctgcagcgcct  
gggtgcgcccaggacatcgacctgacccactcagctcaggtcttggcaatttgccgcg  
gcctttccagctgtgctgcgcccgcgtaacgccccggccttcagattcgctgtga  
40 gcgatgaggcccccacatgtgggttgctacgtcccttcttggtccacagcttgccctca  
tccctctggcagtcacgctgtcttccctgcatgctgtgctgcccgcgtgccttccccca  
tccctccaatctgtgccccgtctgcagggcagaggcagatgtggcatgtcttttgtacc  
cacagagcttgttctatggcacgccttttctaggtttttgtatcatttcttagtacat  
45 tgtaatagattcaaaccagctcttgg

>bc001517 (TOXMARKER Assignment: 59; SEQ ID NO: 145)  
ggcacgaggcaggcttcatttggagtcaggcctggctgttgctcaggtgaccagcttggtg  
tctctgggaggggcgctgctttcccggccaccggcgcatgatccagaatgtcggaat  
50 cacctgcgacggctctctttggaaggagaaattgcagaaccctgggctggtttattatc  
agcagcaatctccaaacaagccccaggggctgtcagatccccacgggctggctgtgtag  
agtggaaatcaccttcagcaagtgtcgccctctggaattctattcgggcttggcctctgtg  
ttctccaaccacacatcccgaagtgcgcttacgtgcggggaacgacagtgccatggca  
gacggcgagggtatccggaacccacggaggtgcagatgagccagctggtgctgcctgc  
cacaccaaccaacgtggtgagctgagcgtcgggcagctgctcaagtggattgacaccacg  
55 gcttgctgtccgaggagaggcacgctggctgcccctgtgtcacagcttccatggatgac  
atctatttgagcacaccattagtgttggaagaagtggatgaatatcaaggccaaggtgaac  
cgggccttcaactccagcatggaggtgggcatccaggtggcctcgaggacctgtgctct  
gagaagcagtggaatgtgtgcaaggccttggccaccttcgtggcccgcgagagatcacc  
aaggtgaagctgaagcagatcacgcccggacagaagaggagaagatggagcacagtggtg  
60 gcggctgagcgccggcgcatgcgccttgtctatgcagacaccatcaaggacctcctggcc  
aactgcgccattcagggcgatctggagagcagagactgtagccgcatgggtgccggctgag

aagacccgtgtggagagtgtggagctggtcctgcctccccacgccaatcaccagggcaac  
 acctttgggggcccagatcatggcctggatggagaatgtggccaccattgcagccaggtga  
 gggcaggggtgtgtgcctctgcctccccctctttctcctcctcctcccccttggtacctc  
 cctctggaggggaaaccccagcttgggggtggcattcaaggcttcagaagcttggtgtt  
 5 ctgaatcagagaaatgaatttttgtgaactgaccattccttgttctactaaaaaagctag  
 catcttttacatgggaaacaccaggtctcttggcctggcactagatcctcccccttgatct  
 ggccctacctgcactccttctagtatctatgttcccttcacatcaagccttctagtatct  
 atgttcgcttcacatcaaaccatttgcgttctctgttcccatcctccactttccagcc  
 10 cctgcctttgctcctgatgtagcctcctgcccgtgcttccccctactcttctttgtctgcta  
 atatcctgcccacttccctccataaaagccatctctgactgttcccttcttctaaggggtga  
 aaattgttttctcctctaacaatctgtttctgtccggggcttgttctaccctaaatatc  
 agggatattttttatagttatggtaactgaccttcaactaattgacactctcacacctccaa  
 gactttgctccttgcgttccctctaccaggagtgcctttcccaacccatgcccctttccag  
 ccagggtgattcctccttattcttttagagcctggcttaaatggccccctcctccagttaa  
 15 cctgtgggagacagtgcataagcaatgctgttttgggcaggcctggctatgagtgacgta  
 agatccttggaggagcctgatggtcagggaaaggctgcctgaaggaagagcacttcagctgg  
 gacttgaatgccaaagtagctttgggtaaggggagggttctggatagtgggaacagcagc  
 taccaaggtgtaaaaagtggaaaggaataagggaagggtttacccaaagccctgctttc  
 ttctgtccccctcaaacttggcttctttccagccatgcataagacctcagtattctaaacta  
 20 tgaatgggactttagtctgtgcctctgggcagaactgccactgggttgggtggcagtg  
 ggtgggtcagaatgtgtagttccaggctgcgtctggggatgggaccaggggtagaaggccg  
 gccaagctggcctagcatggtggctcacacctgtaatcccagcgctttgggaggctgag  
 gccacttaggccagaagttcaagaccagcctgggaacaaggtaaaacctcatctacta  
 aaaatacaaaaaattagccaggtgtggtggtgcgtgcctgtagtctcagctacttgggagg  
 25 ctgaggcaggagagtcacttgaacccgggaggcgagggttgacgtgagccgagattgcac  
 cactgcattccagcctgggcaacagagtgaacccctgtctcaaaaaaaaaaaaaaaaaaaa  
 aaaa  
  
 >bc014075 (TOXMARKER Assignment: 57; SEQ ID NO: 171)  
 ggagcgcagtcgctccgcgatggactcgccgggtcccggcctctatgttcgcccccgagcc  
 cagctccccggggcgccagggccgcggcgccgcgcgcgcgactccacggcggttttga  
 ctcgactgcagcgaggacggcgaggcgctcaacggcgagccagagctggacctcaccag  
 caagctggttctagttagccctacatcagagcagtagacagcctacttcggcagatgtg  
 30 ggagaggatggacgagggatgcggagagaccataatgtcattgggcagggatcagatgg  
 gactgagtagtgggctgagtgaaagctgacatggaggcctcctacgccacagtgaaagcat  
 ggccgaacagatagaggccgatgtcatccttctgcgggaacggcaagaagctggggggccg  
 cgtgcgtgattacctggtccggaaacgagtaggagacaatgacttccctggagggtcaggg  
 agcagtggtgggcaacgtggatgctggcaaaagcacgcttctgggggtcctgacacatgg  
 35 ggagctggacaatggccgaggctttgcccgcagaaactcttccgccacaaacatgaaat  
 tgaatctggtcgcaccagcagtggtgggcaacgacattctgggctttgacagtgaaaggcaa  
 gtccacgaaagtcatctaccttcacgacttggctggtcatgagaagtacctgaaaaccac  
 tgtcttcggcatgacagggccatctgcctgacttctgcatgctcatggtgggcagcaatgc  
 40 tggcatcggtgggcatgaccaaagaacacctgggcttggcactggcactcaatgtacctgt  
 ctttgtggtagtccaagattgacatgtgtcctgccaaacatcctgcaagaaacctgaa  
 gctgttacagcgctgctgaagtaccaggtgcgggaagatccccgtgctggtgcagag  
 caaagatgatgtgattgtcacagcctccaacttcagctctgaaaggatgtgcccgatatt  
 ccagatctccaacgttacaggcgagaacctagatctgctgaagatgttccctcaacctcct  
 45 ctccccccgcaccagctacaggaggaggagcctgctgagtttcagattgatgacacctta  
 ctccgtcccgggtgtggggacagtggtttcggggacaacactgagaggcctgatcaagct  
 gaatgacacgctgctgctggggccagacccttgggtaacttccctgtccattgctgtcaa  
 atccatccatcgcaagcgatgctgtcaaggagggtgcggggtggccagacagcatcctt  
 50 tgcgctgaagaagatcaagcgctcgtccatccggaagggtcagtgatggtttccccacg  
 tttgaatccccaaagcctcctgggagtttgaggccgagattctcgctcctccaccacccac  
 cacaattagcccgctaccaggccatggtgcactgtgggagcatcaggcagacagccac  
 55 cattctgagcatggacaaggactgtctgcgactggggacaaggccactgtacacttccg  
 cttcatcaagaccctgagtacctgcacatagaccagcggtggtgttccgggaaggccg  
 caccaaggctgtcggcaccatccaagctcctccagaccaccaacaactccccaatgaa  
 ctccaagccgcagcagattaaaatgcagtcgacgaaaaagggtccccctgacgaaacgaga  
 60 cgagggggggcccgctggtgggcccagcagtaggagcaccacccctggagatgaagcctc  
 ctctgtaggggcagggcaaccagctgcgtccagcaatctccagcctcagcctaagcccag

5 cagtggagggccggcgacgagggggccagcgccacaaggtgaagtcccagggggcctgtgt  
 gactcctgccagcggtgctgaaccttccccctggcccacctcaccacccaaggggtcat  
 catctctggccaccactccaccagatgggcagagcagctatgaccgccacccagccctcc  
 cgctcaggccacagccggagctccgcattgccccacccccattttccaggggggtgtg  
 aatttataagctgacgaaggttagccagacttccggaggactgacctctctcactgtcct  
 cccaccttcttctcactcacacattttttgtacatctgggcccttagtttttattctg  
 tttattatatgtctctgtctctctctattgtgtgtgtgtgtgtgtgtgtgtgtgtgtg  
 tgtgtgtgtgggtgcaggagtgccacccccagggccctgtcaacctctcttttctcctcca  
 10 tggctgtctgcctgcgtatctgtctctgagaatcctcggggcggtcaggggatgtcagga  
 ggggaaggagccgcccctccctatcttgctgctcctcttgccactcaggggcaccttccat  
 ggagccagaccgggtggaggggcttctgggatttgggtgtctgctgctgccagagcaggaa  
 cccccagtctaggacttgggcatttttaacaggagaaagttagtggttcccttttctctc  
 tctcctcctttttccctttaagcccacagattcaggtcatgccaaaagctctctggttgt  
 aacctggagacatgtggaggggaatggcgatgggattataggactctcccatctcgggc  
 15 cctgacctgaccttggccaccaacccaaagacagctggtgggtttcccttggagacaa  
 tccctgcgtttgcctggggccggccctggctgcccctcagcttctcgctgatctgccggcctg  
 gagcctcccatcaccccgcttcttgttgggcctcaggcactggttaccagaagggggtct  
 gggctctgctcaggatcatgtttttagcacctcctgttggaggggtggagggatgttccc  
 ctgagccaggctgagactagaaccccatcttccctgagccaggctgagactagaacccca  
 20 tcttccccaccacgccacccctgtggctgctacaggagcacagtagtgaaggcctgagct  
 ccaggtttgaaagacccaactggagcgtggggcgggcaggcaggggttagtgaaaggaca  
 cttccagggttaggacagagcatttagccttctggaagaacccctgcctgggggtgggact  
 gtgcaggccagagaaggtggcatgggcctgaacccacctggactgacttctgcactgaag  
 ccacagatggagggtaggctggtgggtgggggtggttcgttctctagccggggcagacac  
 25 cagctggctgggtccttccctcagccttgccctcctctgtccccaaccccttcccttccct  
 cctgcttgcggactgtggtccctcctccttccctccttccagctgtttctagtaccac  
 ctacccctggcgtggactgatcagaccagcattcaaaaataaaagtttgttccaaaaaaa  
 aaaaaaaaaaaaaaaaaaaaaaaaaa

30 >ak027871 (TOXMARKER Assignment: 122; SEQ ID NO: 134)  
 gtttgacgatgagagtgatggggaagaagaggaggagctcatggatgaggatgtggaaga  
 agaggatgactcagagatctcagggtacagcgtggagaatgccttcttcgatgagaagga  
 agacacctgtgctgccgtgggggagatctctgtgaacaccagtgtggccttccctccata  
 35 catggaaagtgtccttgaagaagtatttaaactgctggagtgcctcacctgaatgtgcg  
 gaaggcagcccatgaggctctgggtcagttttgctgtgactgcacaaggcctgtcaaag  
 ctgccctcggaacccaacactgctgcttgcaggctgccttggcccgagtcgtgccatc  
 ctacatgcaggcagtgaacaggagcggaacgcagggtggtgatggccgtgctggaggc  
 cctgacaggggtgctccgcagctgtgggacctcacactgaagccccctgggcgcctcgc  
 40 tgagctctgtggcgtgctcaaggctgtgctgcagaggaagacagcctgtcaggatactga  
 cgaggaggaggaagaggaagatgatgatcaggctgaatacgacgccatgttgctggagca  
 cgctggagaggccatccctgccctggcagccggcgtgggggagactcctttgccccatt  
 ctttgccggtttcctgccattattgggtgtgcaagacaaaacagggctgcacagtggcaga  
 gaagtcccttgagtgaggaccttggcagagactattcagggcctgggtgctgcctcagc  
 45 ccagtttgtgtctcggtgctccctgtgctgttgagcaccgccaagaggcagacccga  
 ggtgcgaagcaatgccatcttcgggatgggcgtgctggcagagcatgggggccacctgc  
 ccaggaacacttcccaagctgctggagctccttttccccctcctggcgcgaggcagaca  
 tgatcgtgtccgtgacaacatctgtggggcacttgcccgctgttgatggccagtcaccac  
 caggaaccagagccccaggtgctggctgccctactgcatgccctgccactgaaggagga  
 50 cttggaggagtgggtcaccattgggcgcctcttcagcttccctgtaccagagcagccctga  
 ccaggttatagatgtggctcccagcttctgcgtatctgcagcctcattctggctgacaa  
 caagatcccaccagacaccaaggccgactgttgctgctcctgacgttccctggccaaaca  
 gcacaccgacagctttcaagcagctctgggctcactgcctgttgacaaggctcaggagct  
 ccaggtgtactgggcctctcctagactgcaggctgcagccagtccagagagaatagagc  
 55 ctgcccaggccttaagaccacctctcagcccagttcagttctgccttaccaaagattctg  
 agactcataccatttggagccagccccacttgctgccttacagggtgtccctgaggct  
 ggtctgttacaatatgagtcagatcatcactgttaataaaagcagcttgttttctgctt  
 gaacaatag

60 >bc003394 (TOXMARKER Assignment: 63; SEQ ID NO: 149)  
 cgaagactgagcggttgtggccgcgttggcgacctccagcagcagtcggcttctctacgc  
 agaacccgggagtaggagactcagaatcgaatctcttctccctcccttcttgtgagatt

tttttgatcttcagctacattttcggctttgtgagaaaccttaccatcaaacacgatggc  
 cagcaacgttaccaacaagacagatcctcgctccatgaactcccgtgtattcattgggaa  
 tctcaacactcttgggtcaagaaatctgatgtggaggcaatcttttcgaagtatggcaa  
 aattgtgggctgctctgttcataagggtcttgccttcgttcagtatgttaatgagagaaa  
 5 tgcctgggctgctgtagcaggagaggatggcagaatgattgctggccagggttttagatat  
 taacctggctgcagagccaaaagtgaaccgaggaaaagcagggtgtgaaacgatctgcagc  
 ggagatgtacggctcctcttttgacttgactatgactttcaacgggactattatgatag  
 gatgtacagttaccagcacgtgtacctcctcctcctcctattgctcgggctgtagtgcc  
 10 ctgaaacgtcagcgtgtatcaggaaacacttcacgaaggggcaaaagtggcttcaattc  
 taagagtggacagcgggatcttccaagtctggaaagtgtgaaaggagatgaccttcaggc  
 cattaagaaggagctgacccagataaaaacaaaaagtggattctctcctggaaaacctgga  
 aaaaattgaaaaggaacagagcaaaacagcagtagagatgaagaatgataagtgcagaaga  
 ggagcagagcagcagctccgtgaagaaagatgagactaatgtgaagatggagtctgaggg  
 15 ggggtgcagatgactcgtctgaggaggggacactactggatgatgatgataatgaagatcg  
 gggggatgaccagctggagttgatcaaggatgatgaaaaagaggctgaggaaaggagagga  
 tgacagagacagcgccaatggcgaggatgactcttaagcacatagtgggggttagaaatc  
 ttatcccatattttctttacctaggcgcttgtctaagatcaaatttttcaccagatcctc  
 tcccctagatcttcagcacatgctcactgttctcccatccttgctccttcccatgttca  
 20 ttaattcatattgccccgcgcctagtcccattttcaactcctttgacgctcctagttagtt  
 ttgtaaagtcttacctgtaatttttgcttttaattttgataacctctttatgacttaaca  
 ataaaaaggatgtatggtttttatcaactgtctccaaaataatctcttggtatgcaggga  
 gtacagttcttttcatcacaataagttcagtagttgcttccctaactgcaaaggcaat  
 ctcatattagttgagtagctcttgaaagcagctttgagttagaagtatgtgtgttacacc  
 25 tcacattagtgtctgtgtggggcagttcaacacaaatgtaacaatgtattttgtgaaat  
 gagagtggcatgtcaaatgactcctctagaaaaataattagtgttatagtcttaagatt  
 tgttttctaaagtgtactgtgggttattttgtgaacagcctgatgtttgggaccttt  
 tttcctcaaaaataaacaagtccttattaaaccaggaatttgagaaaaaaaaaaaaaaaa  
  
 >bc006191 (TOXMARKER Assignment: 79; SEQ ID NO: 156)  
 30 ggcacgaggctttccccccccccccggccattaccgaagcggatgaaaacaaacacta  
 acgatggcgccgcccgggaagcgaccggctgctgggcttaaggcgggagtgcacgcttaac  
 cagtggggaagcactgaagagcgccagtcgacgtgggtgcgacaactcgcgaggtctta  
 ggagcaaaacgtctggggcctgcagccaggacccttctgaagccttaggtgtctatcgg  
 35 cgacgtgtacggtcactgacgtcctcgagcgcggaaccctcagccaggagcgcggtgg  
 tcgggtcccagggtcccggcctccgtaatgagagcccgaaccactctttgtgccgcagctt  
 cgcagcatcttgactcaagtgatttctcctgcctcagcctcctgagtagctgggactaca  
 gattcctataggcaatggaaactgatctcaattcccaggacagaaaggacctggacaagt  
 40 ttattaaatttttggccctcaagactgtccaagtattgtccaggctcggcttggtgaaa  
 agatttgcactcgttcatcatcttctccaacgggttcagattgggttcaacttagcaatca  
 aagacatcccagagggttacacatgaagcaagaaggcactggcaggacagctgcctgcag  
 tcgggagggtccatgtgtgtggagatttcacttaagacttctgagggagattccatggagc  
 tggaaatatggtgtcttgaaatgaatgaaaagtgtgataaagaaatcaaagtttcctaca  
 45 cgggtacaaacagactgtcattgctgctgaagtccttcttgctataactagggtgacac  
 cagcctataggctctccaggaaacaagggcatgaatatgtcatattatacaggatatatt  
 ttggagaagttcagctgagtggttaggagaaggcttccagacagttcgtgttgggacag  
 tgggcacccctgtgggcaccatcactctttctgtgcttacagaattaacttggcattca  
 50 tgtctaccaggcaatttgagaggaccccacctatcatggggattattattgatcactttg  
 tggaccgtccctatcccagctcctctcccatgcacccctgcaattacagaactgctgggtg  
 aggacactggagtaatataccgtctgtagaagactctcaagaagtgtgtaccacctctt  
 tttccacctccccaccatcccagctgatgggtcctgggaagggaagggtggggtaccccttg  
 ctcccaaccagcctgtccatggtaccaggctgaccaggagagactggcaacctgcaccc  
 55 ctctgacagaacccactgtgctgccacacccctccagtagtgaggatactgaaacccgtat  
 caaacgcagtgagggaacgggctccctcacgatgtctggagaccatctttgtccgaa  
 aagtgggggcttttgtcaacaaacccattaaaccagggtgacctgacgagtttggatatac  
 cctttgccatgtttgtcccaagaatttggagctggaggataccgatccaatggtgaatc  
 ctccagattcccagagactgaatctcctctccagggcagcctgccttgcagctggcccc  
 60 ttccctgctgctgtcaccatccactgtttgacattccagctgggtggccaagagattggt  
 gtggaggcagaaaagggaaggagacagtgccaggaggaagaagggaaggagtcccttagct  
 ctcttcattgtcccttttacttctgctatcttcttctcctcttcttctctcttgcct  
 ctatgcctgtatttctggcaatatgacaggcctgcctacccaagatcagaactccaaac  
 cactcccacccctgaagggtcgggagggtctgagcagccctgggtggctgctgtcagg

5 tctctcagctccatgggaaataaaaaatggcaccctgaatctctaggattttgtcacttggg  
 gtcacagcaaagttctcttctcttgtcccccgttgtgctccttgggtatagaacatg  
 gtaaatatatttactttcagagaaaccagatattttatagaggaaatatgtttgaggtg  
 agttgtttttcacttggagaaggcggagggctcttcttgggacggagacctcctcctccg  
 10 gaggttattgagaatccgggctgctgctttgaggatcttcccaccatacagacagcgaga  
 tccaagaagagggtggccgggggcaaagtcacctcccagtggtgactggaactga  
 ctaaaggctttaccttggatagttgcgtattcctgggtgagagccttacatctcccacagt  
 ttctgcagagtgaactgactccattctggcagcccaggaagtctgggtgctaaatgtgat  
 ggccacatgtagtgggttaggggatgttgtgtgtgtccccaactgctgggtacttgttc  
 ctgatccctggggctgtcctgtggagcttttctcctgcttgggcctagctaccatctcc  
 ctctaataccaggttctctacactgccctgggggttaccagctggattggcttctgggtg  
 agaaatcaaagctgggcgtatgattgacttaacccttcaggatattgttacttgaataagt  
 caagtgcctagcctcaccacacctatgatctgtcctttcccagcctcgctggtagtcttg  
 15 tcaaggagatctagggtctactccattcctcctggcccacctggggcattcactggcagca  
 gctgtgcttcagtgagcaggtgggtctcagctgcttggttagtatactgcatgtgacact  
 gttcccacatacaaggctgacttctgaggattggagcaggtctggcggggaccagagct  
 ctgctgctgctgctgccaccaagaagtgttagcagaagcagtagcagccaactggcct  
 cctgactttggcccagagcacatgcgtggcttgtgaaccacaggtcaggtttatcccca  
 agggcccagctttgagaagggggaaggccctggtaagttattgatgcccccatatttca  
 20 gctactgctctctttccaaggccttgcattggaaggcctagccattgtctgaggcagcaa  
 tctttggcatctacaggtggcagcagcctttcaccagggtccatctgtgaagagtctca  
 gccatgactttgagctgagcttgggagaagtaaaagcaactgttaaggccagcccttgccc  
 ctacagacctgccatgaaaggaatgagccctagactgactcctgcagcaccctgggacag  
 gctgggaccagctgtttgtctccaggtgtcagagtcctcctcctcctccaacctctcca  
 25 acctactttgtttggaaataccgagctacacttcaaatgtattcaagggtttccaata  
 aatttttttctgtataaaaaaaaaaaaaaaaaaaaaa

>ab014606 (TOXMARKER Assignment: 53; SEQ ID NO: 143)  
 30 cgctccccccccagctcgctgcccgggcccggcgccgctggcgccgctactgctg  
 ccgccccggggcgcgagtcgcccggccgcccgggacccggcgaggggcccggggca  
 gctccgaaccggccccagatccttcccgttccgcctcacgcttcccggaaagcttgctc  
 ctctccgcccagctgctccgggagccccgcccgcgcgaggggtatctcccagagccccagc  
 tgggtgtggccaggccccaggagtaggatggggctccccctacgagggccgggtggcagcca  
 35 gaactgatacagccccctggctcggggccaggagccagctgaggagggcaggagtgtc  
 tggagctatggctggtgctcggtagaagtgccagtgaggggttcggccctttaacgccc  
 tgagaccagccaggatgccaagtgtgtggtcagcatgcagggcaacaccacctccatcat  
 caatcctaaacagagcaaggatgcccccaaaagcttcacctttagactactcctactggctc  
 acacacttcgacggaggacccccagtttgcatctcagcagcaagtgtatcgggacattgg  
 agaagagatgctgctccacgcctttgaaggctacaacgtgtgcatctttgcctatgggca  
 40 gaccggggctgggaaatcctataccatgatggggcgacaggagccaggggcagcaggggcat  
 cgtgccccagctctgtgaggacctcttctctcgcttagtgagaaccagagtgtcagct  
 atcctactctgtggaggtgagctatatggagatctactgtgagcgggtacgagacctctt  
 gaaccccaagagtcgggggttctctgcgggtccgggagcaccatcctgggcccgtacgt  
 gcaggacctgtccaaattggctgtgacctcctacgcagacattgctgacctcatggactg  
 45 tggaaataaagcacggactgtggctgccaccaacatgaatgagaccagcagccgttccca  
 tgccgtctttaccatcgtcttcacacagcgtgccatgaccagctcacggggctggactc  
 ggagaaggtcagtaagatcagtttgggtggaccttgctgggagtgagcgagccgactcctc  
 agggggcccgggcatgcgcctgaaggaaggagccaacatcaataagtccctgactacact  
 agggaaagtgatctcgcccttgacagatgcaatcaaagaagcgaaagtccgattttat  
 50 cccctacagggaactctgtgctcacctggctgctcaaggaaaaatttgggggggaactcacg  
 cacagccatgattgcagccctgagccctgctgacatcaattacgaggagactctcagcac  
 cctcaggtatgctgaccgcaccaagcaaatccgctgcaatgccatcatcaacgaggacc  
 taatgcccggctgattagagagctgcaggaggaagtagcccggtgcgggaactgctgat  
 ggctcaggagctgtcagcctctgctctggaaggcctgaagacggaagaggaggtgtcag  
 55 aggcgccttgccagctgtgtcatctccccagctccagtttaccctcatcaccaccac  
 acataatggggagctggagccgtcattctcccccaacacggagtcacagattgggcctga  
 ggaagccatggagaggctgcaggagacagagaagattatagctgagctgaacgagacatg  
 ggaggagaagctacgcaagacagaagccctgaggatggagagagaagcattgctggctga  
 gatgggggtggccgtccgggaggatgggggaactgtgggcgtcttctctccaaagaagac  
 60 tccccacctgggtgaacctgaacgaagacctctgatgtctgagtgctgctctaccacat  
 caaagatggcgtcaccagggtcggccaagtagatatggacatcaagctgaccggacagtt

cattcgggagcaaacactgtctgttccggagcatccccagccagatggagaagtgggtggt  
 cactctggagccttgtgaaggagctgagacatatgtgaatgggaagcttgtgacggagcc  
 gctggtgctgaagtcagggaataggattgtgatgggcaagaaccacgttttccgcttcaa  
 5 ccacccgagcaggcaaggctggaacgggaacgaggggtccccccacccccaggaccgcc  
 ctctgagccagtcgactggaactttgcccagaaggaactgctggagcagcaaggcatcga  
 cataaaagctggaaatggagaagaggctgcaggatctggagaatcagtaccggaaagaaa  
 ggaagaagccgatcttctgctggagcagcagcgactgtatgcagactcggacagcgggga  
 tgactctgacaagcgctcttgtgaagagagctggaggctcatctcctccttgcgggagca  
 10 gctgccgcccaccacggtccagaccattgtcaaacgctgtggtctgccagcagtggcaa  
 gcgcaggggccctcgcagggtttatcagatccccagcgcagcaggctgcaggggcaaaga  
 cccccgctggggccaccatggctgacctgaagatgcaggcggtgaaggagatctgctacga  
 ggtggccctggctgacttccgccacgggccccggtgagattgaggccctggccgcccctcaa  
 gatgcgggagctgtgtcgcacctatggcaagccagacggccccggagacgcctggagggc  
 15 tgtggcccggtatgtctgggacactgtaggcgaggaggaaggaggtggagctggcagtgg  
 ttggtggcagtgaggagggagccccgaggggaggaggtggaggacctccgggcccacatcga  
 caagctgacggggtattctgcaggaggtgaagctgcagaacagcagcaaggaccggagct  
 gcaggccctgcgggaccgcatgctccgcatggagaggggtcatccccctggccccaggatca  
 tgaggatgagaatgaagaaggtgggtgaggtccccctgggccccgcctgaaggatcagaggc  
 20 agcagaggaggcagccccagtgaccgcatgccgtcagccccggccccctcgcgccact  
 gtcaagctgggagcgggtgtcacggctcatggaggaggacctgccttccgtcgtggtcg  
 tcttcgctggctcaagcaggagcagctacggctgcagggactgcagggctctggggggccg  
 gggcggggggctgcgcaggccccagcccgccttctgtgccccctcacgactgcaagctacg  
 cttcccccttcaagagcaacccccagcacccgggagtcttggccagggtgggggagcgggga  
 25 ggctccaactccgctccaacccccctgaggaggtcactccccatccagccacccccctgcccg  
 ccggcctccgagtccccgaagggtccaccatccccgcaggaaactccttgatggaggggg  
 ccgatccccgggagcgggttctgcacagcctgaacccccagcacttccagccccaaaagca  
 caactcttatccccagccaccccccaacctacccagccccagcggccccccaggccccgcta  
 ccccccatactactccccacgaatgagacggcagcgcttctgccccctgacctcaagga  
 30 gagtggggcagctgtgtgagtgccacatcctgggcagaggccctgggtggggcccccttgc  
 aggagaagggaagacgcccagacgctgcttccccagaagtgtggtggggcagggaggccca  
 ggagatgagagagaaggtccgagttaggtgatagaagacaagggggagaccgagccggagg  
 ctgaggaaagggaagaggggcacggagtggccaggagcaaaccaaagtgaagagagagatag  
 gaagctgcctcggggccaccccttgcaaaggggggtgtgtccacaaacgctgctatgggt  
 35 ggggtgggggctgggtgctgcgtagccagtgtttgactttcttttcaagtgggggaaa  
 gtgggagaggactgagagtgaggcaagttctccccagccccctgtccgtctgtctgtctgt  
 ctgtggtggtttctgtttcttgggaggcatggttaggatcataagtcattccccctcccc  
 ccaggcctcctgctatatatttgggggacctgactgggttggctggagtcccatgaggatgt  
 gggccctttaataaaggatagcaaacaggggagcttgtggcctgtttgttttgggttttca  
 40 tggaggtgtaggttatataaggcaatggcacaggctcttaagcatacttatcagtgaagta  
 ttgtatgtgtgctctgtgcaggcaccacccagatctggatataagaatgtttccatcttg  
 tcttctgaacttcacccctcctgtctcttcccttcagggttgcgcagcccgatcttttcccc  
 gcttttttttttgggagacaggggtcttgctttgttgccaggctggagggtacagtcttg  
 gctcactgcagcctccgcctcctgagtagctgggattacaggcatgtgccaccacgcccg  
 45 gctcattactgtttttttgtagtgcagaggtttcgccgtgttggccaggctgggtctcga  
 actcctgatgacctcaagtgatccgcccgccttggcctcccagagtgggtgggattgcaga  
 gacagtgatcttgctatgctgcccaggctgggtctcaaccttatgggctcaagtgatcctc  
 ccacctcagcctcccaaagtgttaggattacctgcgtgagctacagcgcctgctgttct  
 tgggcttcttgcagagcctcttcagctgcagagaagcagctctcctttctccaagtcag  
 50 agccaacaggacgaataatgaagctgttgggaagatttactgataatacatgtaaagggt  
 ctagcacatttttaggagctcaagggtgggtgccttcccttttttctttactctgaaccggat  
 atgaggccttgagaaagaagagagggcgcttgcaaacgaggtgaggtctcaggcacagtg  
 gctcacgcctgtaatcccagcactttaggagaccgaggcgggcggtatcatgaggtcagga  
 gttcgagaccagcctggccaacatgggtgaaagcccgtctctactaaaaatacaaaaatta  
 55 gacgggcatgggtgggtgggcacctgtaatcccggctacttgggaggttgagggaggagaat  
 cgcttgaacccaggaggtggaggttgcagtgcagccgagactgcaccattgcactccagcc  
 tgggcaatagagcgagactccgtctcaagcaagcaagcaagcaaacaaaataaaaa  
 acgaggtcaagtttcaaaagatgtcacccccaaacctggcaaaacttctcctcaagcctg  
 tcgttccactcttgtccgccaggaggagaaaaggttccctcgaaggacgtctttgcttgc  
 60 gcgttcacggagccttgagaacgagtgggcgaggggaacccctgcggccctgcgcgcctaa  
 gggaggacctgactcctttcagaagtagcatttcttcccccttcgtgggtgctcttgagtt  
 ccaaagaaaaggagaagagccttcattgagcagcttcttctgccttagggactgtgcta



5 ggggtagatcgaccttaggggaaacaatccccgcttattagaggaggttttgatcagg  
 gtttgctttatttgaatttaacaaatacagaaaagcagaaggaagaaaattgaagtaat  
 ccatgtttccactgggcgcgcggtcacgcctataatcccagcactttgggaggccaag  
 gcgggaggatcacgaggtcaggagttcgagaccagcctgaccaacatggtgaaaccccc  
 gtctctactaaaattacaaaattagccggtcgtggtggcacacgcctgtaatcccagcta  
 ctcaggaggtgaggcaggagaatcgcttgaacccgggaggcagaggttgacgtgagccg  
 agattgcaccactgcactccagcctgagcaacagagtgagactccgttgtctttaaggc  
 caatcccatagcaaatacagagactcacttgagtaagaaagggttttgacaagaaaacc  
 10 cacagaagaagaggtgaagctgtggatataagaaaggcactacaatctgtatttaaatga  
 attgcacactagataatatgtatgaaaaattattt

>u78525 (TOXMARKER Assignment: 61; SEQ ID NO: 148)  
 aattccggccgctcgacgggagagtcggtagcgcgggcgaggccctgcgagtag  
 15 gcaagcgttgggcccattgcaggacgaggagaacgtggcggtgcccaggcgccgaggag  
 cgcgccgagccccggccagcagcagccggcgccgagccgcccagccgaggggtgctg  
 cggcccgcgggcccgcgctccggaggccgcggggaccgaggcctccagtgaggaggtg  
 gggatcgcgaggcgggcgggcggagcccagggtgaggaccgagccggcgaggcgagag  
 gcgccctccggcccgctccgagtcgcccctcgccgcccggcgccgaggagctgcccgggtcg  
 20 catgctgagccccctgtccggcacagggcgaggccccaggagagcaggctcgggacgag  
 cgctccgacagccgggcccaggcggtgtccgaggacgcccggaggaaacgagggcagagcg  
 gccgaggccgaacccccggcgctggagaacggcgacgaggacgagccctcctcagcgac  
 cccgaggacttcgtggacgacgtgagcgaggaagaattactgggagatgtactcaaagat  
 cggccccaggaagcagatggaatcgattcggtgattgtagtggacaatgtccctcaggtg  
 ggacccgaccgacttgagaaactcaaaaatgtcatccacaagatcttttccaagtgtggg  
 25 aaaatcacaaatgatttttatcctgaagaggatgggaagacaaaagggtatattttctg  
 gactacgcgtccccctgccacgctgtggatgctgtgaagaacgcccagcggtacaagctt  
 gacaagcagcacacattccgggtcaacctctttacggattttgacaagtatatgacgac  
 agtgacgagtgggatattccagagaaacagcctttcaaagacctggggaacttacgttac  
 tggcttgaagaggcagaatgcagagatcagtacagtgatgttttgagagtggagaccgc  
 30 acttccatattctggaatgacgtaaaagacctgtctcaattgaagaaagagcgagatgg  
 acagagacgtatgtgcgttggtctcctaagggcacctacctggctacctttcatcaaaga  
 ggcattgctctatgggggggagagaaattcaagcaaattcagagattcagccaccaaggg  
 gttcagcttattgacttctcacctgtgaaaggtaacctgggtgacctttagccccctgatg  
 gacacgcaggatgacctcaggccataatcatctgggacatccttacggggcacaagaag  
 35 aggggttttactgtgagagctcagcccattggcctatttttaagtggagccatgatggc  
 aattctggaatgggaatgacctggatcgttagcatctatgaaactccttctatgggt  
 cttttggacaagaagagtttgaagatctctgggataaaaagacttttcttggtctcctggt  
 ggtaacataatcgcttctggtgctgaagacaaagatattccagccagggttaaccttg  
 atgcagctccctaccaggcaagagatccgagtgaggaaacctgttcaatgtggtggactgc  
 40 aagctccattggcagaagaacggagactacttgtgtgtgaaagtagataggactccgaaa  
 ggcacccagggtgtgtgcacaaattttgaaattttccgaatgagggagaaacaggtacct  
 gtggatgtggtcgagatgaaagaaaccatcatagcctttgcctgggaaccaaaggaagt  
 aagtttgcgtgtgctgcacggagaggctccgaggatctgtgtctttctaccacgtcaaaa  
 aacaacgggaagattgaactcatcaagatgttcgacaagcagcaggcgaacaccatcttc  
 45 tggagcccccaaggacagtctgtggtgttgccggcctgaggagtatgaacggtgcctta  
 gcgtttgtggacacttcggactgcacggtcatgaacatcgcagagcactacatggcttcc  
 gacgtcgaatgggacatcctactggcgctacgtcgctcacctctgtgtcctgggtggagccat  
 aagggtggacaacgcgtactggctgtggactttccaggggacgcctcctgcagaagaacaac  
 aaggaccgcttctgccagctgctgtggcgccccggcctccacactcctgagccaggaa  
 50 cagatcaagcaaattaaaaaggatctgaagaaatactctaagatctttgaacagaaggat  
 cgtttgagtcagtcacaaagcctcaaaggaattggtggagagaaggcgaccatgatggaa  
 gatttccggaagtaccggaataatggccaggagctctatatggagcagaaaaacgagcgc  
 ctggagtgcgaggaggggtggacactgacgagctggacagcaacgtggacgactgggaa  
 gaggagaccattgagttcttcgtcactgaagaaatcattccccctcgggaaatcaggagtga  
 55 cctggagcactgtggggacggactccgcctgctgttcccgcgctgagctacaggactccc  
 gagtgtgagccgcggttctctgttgcagcgcagccgtgtgtgctgtggagccgagggccg  
 tcctgcaggaaagccgctgactccgcctcctccctgtgctctctggctctggactgtga  
 ctgcgcctggattctgccattgcgacacatttttgtgcctttcagccccctgggtgtgtga  
 gtgggggatttaaggcaccgccttccacttcttctgtttggagtgttctgttggaaacc  
 60 gccggcgttggtccgaagacttagcgacgccactggcgccacttctcctgcgcccagt  
 gatgtttccacggtgcctgtacacagccgagcagcatttccgttgaaggacttgcaccc

cattgcgggcagtgctggacgtgtcccgagacccaccgggagggcgccgcatgccttgta  
ccccaccgtgcaggttggtggccggttttctccgcaggttgaacatggaataaaagcaa  
acttgatatggaattcaaa  
aa

5

>bc014348 (TOXMARKER Assignment: 88; SEQ ID NO: 157)  
ggggagtgccgggcggtcggcggtcagggcagcccgggcctgacgccatgtcccggaa  
cctgcgcaccgcgctcattttcggcggttcatctccctgatcggcgcgccttctatcc  
catctacttccggccccctaatagagattggaggagtacaagaaggaacaagctataaatcg  
ggctggaattgttcaagaggatgtgcagccaccaggggttaaaagtgtggtctgatccatt  
tggcaggaaatgagagggtgtcatcagctctgattaagaaaggagatttcttcatgctt  
tcgattctgcattgggtacagccagtcacctcaccagagaatgacggctggagaagaaaa  
ctctgtaataccataaataagagtgttgaataaaaagactgtgcacaaggattaatatt  
tcccttcttaagtatcaaaagaactctggaacaaattataaccattaggaaggttttcatg  
attcagttgattttccaaaaatgaagctatctcaccagctgggtttggaggagcaatct  
gcttattattctgtcggttaccacttactcaagcgagctgtgatatgaatacaagcaacca  
gtgggctcgggaaggtccgggtctcttctgccatcttccagataagagatttcagtaaaa  
aactgccatgctgagctgccttatagagctcttcgaaaatgttcgagttgataaagctct  
ttgaggacaaggtacttctgtgcacctcatgctgaagattgcaccgtgttgaaaaataaat  
atgaagcaagtcaaactagatgcatacacttgtgtagaatcaataatcaattaatagaa  
gtgaaaaaatagacattaaaatgatttttcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa  
aa

10

15

20

>m94151 (TOXMARKER Assignment: 89; SEQ ID NO: 165)  
gggctgcaggaattccccacagagggagcatgacttcggcaacttcacctatcattctg  
aaatgggacccccaaaagtttggaatccggacgctaacagtggaaaggctgttgagcca  
cttggtacacaggtgactacacttgtcaacacaagcaacaaaggcccatctggtaaaaag  
aaaggagggtcaaagaaagcccatgtactagctgcctctgtagagcaagccactcagaat  
ttcctggaaaagggtgaacagatcgctaaggagagtcaagatctcaaagaagagtgggtg  
gctgctgtagaggatgtgcgcaacaagggtgagacgatcgccatcgctcctccgagttt  
gcagatgaccttgctcggttaagcgcgccaccatggtagcgggcggaagggtttg  
ctctccgcggtgacacgcttactcatcctggcgacatggcagatgtcatgagactttta  
tcccatctgaaaattgtggaagaggccctggaaagctgtcaaaaatgctacaaatgagcaa  
gaccttgcaaacctgttttaagagtttggaagaaatggtgaaacttaactatgtagca  
gcaagaagacaacaggagctgaaggatcctcactgtcgggatgagatggcagccgcccga  
ggggctctgaagaagaatgccacaatgctgtacacggcctctcaagcatttctccgccac  
ccagatgtcgcgcgtacgagagccaaccgagattatgtgttcaaacaagtcaggaggcc  
atcgccggcatctccaatgctgctcaagctacctcgcccaactgacgaagccaagggccac  
acgggcatcggcgagctggctgcggctcttaatgagtttgacaataagattatcctggac  
cccagtgagttcagcgaggccaggttcggccgctccctggaggagaggctggagagcatc  
atcagcgcgagcgcgtgatggcgactcctcctgcacgcgagacgaccggcgagaggg  
atcgtggcgagtgcaacgcggtgcggcagcgctccaggacctgctcagcgagtacatg  
aataatactggaaggaaagaaaaaggagatcctctcaacattgagattgataagatgact  
aagaaaacaagagatctaaggagacagcttcggaagcagtgatggatcacatatctgac  
tctttcctggaaaccaatgttcttctgtagttctcattgaggctgcaaagagcggaat  
gaaaagggaagtgaagaatatgcccaagtttccgtgagcatgccaacaaactggtagag  
gttgccaatttgccctgttccatctccaacaatgaagaagggtgaaattagttcggatg  
gcagccaccagattgacagcctgtgtccccaggtcatcaatgccgctctgacactggct  
gcccggccacagagcaaagtgtcaggataacatggacgtcttcaaagaccagtgaggag  
aagcaggtccgagtggtgacagaggccgtggatgacatcacctcagtggtgacttcctc  
tctgtctcagaaaaatcacatcttgaggatgtgaacaagtgtgtgatagccctccaagag  
ggcgatgtggacactctggaccgactgcagggggccatcagggggccgggcagctcgagtc  
atacacatcatcaatgtgagatggagaactatgaagctgggggttatactgagaaggtg  
ttggaagctacaaaattgttttctgaaacagtgatgccacgcttcgctgaacaagtagag  
gttgccattgaagccctgagtgccaacgttctcaaccgtttgaggagaatgagttcatc  
gatgcctctcgcttggtgtatgatggcggttcgggacatcagaaaggctgtgctgatgatc  
aggaccccaagaactagaggatgattctgactttgagcaggaagattatgatgtgcgt  
agagggacaagtgttcagactgaggatgaccagctcattgcagggcgagagcgacgggcc  
atcatggcgcaactaccgcaggaggagaaggcaaaaatagctgagcaggtggagatatcc  
catcaagagaaaaagcaagctggatgcagaagtggccaaatgggacgacagcggcaatgat  
atcattgtactggccaagcagatgtgtatgatcatgatggaaatgacagacttcacaaga

25

30

35

40

45

50

55

60

5 ggcaaaggccattgaaaaatacatctgatgtcattaatgctgccaagaaaattgccgaa  
 gcaggttctcgaatggacaaattagctcgtgctggtgatcagtgctcctgattcagca  
 tghtaagcaggatttattagcctaccttcaacgaattgccttgattgccatcagcttaat  
 atctgcagcaaggtgaaggcagaagtgcagaatctgggaggagagctcattgtgtcaggg  
 acaggagttcagagcactttcactaccttttatgaggtagattgtgatgtcatagatggg  
 10 ggcagggctagtcaactttctaccacctcccaacctgtgctgagggagctccgatcggg  
 agtggaaagcagtgattcctccatgctggacagtgccacatcgcttatccaggcagctaaa  
 aacctgatgaatgctgttgcctcacggtgaaagcatcctatgtggcctcaaccaaatac  
 cagaaggtctatgggacagcagctgtcaactcacctgttgtgtcttggaagatgaaggct  
 ccagagaagaagccccttgtgaagagagaaaagcctgaagaattccagacacgagttcga  
 cgaggttctcagaagaaacacatttcgcctgtacaggctttaagtgaattcaaagcaatg  
 gattccttcttaggacgataggttttaacaagaaagctttttctttctttctttcttct  
 15 ttttctttttaattccatttttgtatgcatacctgccagctcgtatgcctctggcatggg  
 gaaattaagggaacagtgtctgtttgcatgtaagatgagatgagatcaataactactgatc  
 catctgtagcctgggaaggagacaggacattcctgtactaagggtggcacagagctgtcct  
 ttgcaacattctcataaaattgggcacagagttcgcattggcgcaatatttatgggagtg  
 ggagggatggggaaaataaacttaactctacaaaagcaaactctaattgcatgcaagaatc  
 attaggttggcaggtatatgcataagtgaaaaatctggaagtgaatggtagaacataaaa  
 acttgtattgcttctgtttcagtgcaaaaatgtactagccaatacgtttaagtgtgtggc  
 20 ccatgaattgaacaatttaaccttgaagtctatatccgtgatattatgtcgatttttaac  
 tgaggggaaattaactagtccagcctaaaatgcttcttttaactctgcattctgtttcctc  
 ttctagtgtgcccattactagtatcatgtttttttccccctttaatgaaaacaataaaa  
 catctatttgagacaattaaaactccttctggggcactggaagcacaatacggtgaccaa  
 tcttgctttcttttttttttttttttaattgaaccatgattttgtagaaatagaaggc  
 25 ccagtggtggaatattagagggaaggaaactgacaacgtgtgaaagtta

>nm\_000521 (TOXMARKER Assignment: 38; SEQ ID NO: 130)  
 ctgatccgggcccgggcccgaagtcgggtcccagggtccggctcggcagaccgggcccga  
 30 agcagccgagcggccatggagctgtgcgggctggggctgccccggccgcccattgctgctg  
 gcgctgctgttggcgacactgctggcggcgatgttggcgctgctgactcaggtggcgctg  
 gtggtgcaggtggcggaggcggctcgggccccgagcgtctcgcccaagccgggcccggcg  
 ctgtggccccctgcgctctcgggtgaagatgacccgaacctgctgcactctcgccccggag  
 aacttctacatcagccacagcccaattccacggcgggccccctcctgcaccctgctggag  
 35 gaagcgtttcgacgatcatcatggctatatatttttggtttctacaagtggcatcatgaacct  
 gctgaattccaggctaaaaccagggttcagcaacttcttgtctcaatcacccttcagtca  
 gagtgtgatgctttccccaacatatcttcagatgagtcttatactttacttgtgaaagaa  
 ccagtggtgtccttaaggccaacagagtttggggagcattacgaggttttagagaccttt  
 agccagttagtttatcaagattccttatggaactttcaccatcaatgaatccaccattatt  
 gattctccaagggttttctcacagaggaatttttgattgatacatccagacattatctgcca  
 40 gtttaagattattcttaaaactctggatgccatggccttttaataagttaattgttctcac  
 tggcacatagttgatgaccagctcttcccatatcagagcatcacttttctgagtttaagc  
 aataaaggagctattctttgtctcatgtttatacaccaaattgatgtccgtatggtgatt  
 gaatatggcagattacgaggaattcgagtcctgcccagaatttgataccctgggcataca  
 ctatcttggggaaaagggtcagaaagacctcctgactccatgttacagtagacaaaacaag  
 45 ttggactcttttggacctataaaccctactctgaatacaacatacagcttcttactaca  
 tttttcaaagaaattagttaggtgtttccagatcaattcattcatttgggaggagatgaa  
 gtggaatttaaatgttgggaatcaaatccaaaaattcaagatttcatgaggcaaaaaggc  
 tttggcacagattttaagaaactagaatctttctacattcaaaagggttttggaattatt  
 gcaaccataaacaagggatccattgtctggcaggaggtttttgatgataaagcaaagcctt  
 50 gcgcccgggcacaaatagttgaagtatggaaagacagcgcatatcctgaggaactcagtaga  
 gtcacagcatctggcttccctgtaatcctttctgctccttggtagcttagatttgattagc  
 tatggacaagatttgaggaaataactataaagtggaaacctcttgattttggcggtagctcag  
 aaacagaaacaacttttctattggtgggagagcttgtctatggggagaaatgtggtgca  
 actaacctcactccaagattatggcctcgggcaagtgtgttgggtgagagactctggagt  
 55 tccaaagatgtcagagatatggatgacgcctatgacagactgacaaggcaccgctgcagg  
 atggtcgaacgtggaatagctgcacaacctctttatgctggatattgtaaccatgagaac  
 atgtaaaaaatggaggggaaaaaggccacagcaatctgtactacaatacaactttattttg  
 aatcatgtaaaaataagatattagacttttttgaataaaaatatttttattgattgaa

60 >ab007887 (TOXMARKER Assignment: 91; SEQ ID NO: 158)  
 gtcagatcagggatcatttttttcttctactccctccccctaccgccccctccc

5      tccctgtttcccttccctccctccctccctctctgctgggtctgtgcgctggggcgccc  
 gatccctccgcagctgggacgctccgaactcgaggcaggagtcggctctccggagcctc  
 gtccttcccttcccttccctgcccccttccccacccccgactcgggcttggcgcgccg  
 gccagaggaaccccgagtcgccggcccaggccccctgagctggagggatggaaaactcctc  
 10      gcagcatcagcctcctcggaggcaggagcagccgctcccaggagatcgaggagctggag  
 cgcttcacgcagcctacgtgctggagtaccaggtgcaggggctgctggctgacaagacg  
 gaggggtgatggcgagagcagaggaccagtcacacatctccagtgagacagcggactgc  
 agcgaaccgctggacagcagctgttcttctcccaggggcgagccccccacagcagaat  
 ggcagcaaagacaactctctggacatgctgggcacggacatctgggaggccaacaccttc  
 15      gattccttcagtggtgccacctgggacctgcagccgaaaagctggacttcacccagttc  
 caccgcaaagtccgacacacgcccgaagcagccccctgccacacatcgaccgcgaagggtgt  
 ggcaaagggaagctggaagatggggatggcatcaacctgaatgacatcgagaaggctcctt  
 ccagcttggcagggtaccacccgatgccccatgaagtggagatcgcacacaccaagaag  
 20      ctgttccgcaggagagaaaatgatcgaaggcggcagcagagacctccggggggcaacaag  
 ccccaacagcatggtgaccaccagccaggcagtgccaaacacaacagggaccaccagaaa  
 tcctaccaggggggctcagcaccacccccctcaggaggcccaactcaccatggctacagc  
 cagaaccggcgctggcaccatggcaacatgaagcaccaccaggcgacaagggggaggca  
 ggcgcacaccgcaatgccaaagagaccatgaccatcgagaacccaaaactggaggacact  
 25      gcaggggacaccgggcacagcagcctcagggcccccccgagccctgacaccctggccccg  
 gtggcttctgagcggctgccccacagcagtcaggggggcccagaggttgagacaaaacgt  
 aaagacagtattcttcccagcgcacgaggagcggcccaaaattaccctgctccagctct  
 tccaaagacagactgcccgaaggctaaaggaaaaggatgaagtggcgtggagacgacc  
 actccccagcagaacaagatggacaagctgatcgagatcctgaacagcatgcggaacaac  
 30      agcagtcagctggacaccaagctcaccaccttcatggaggaggccagaactccaccaac  
 tccgaggagatgctggcgagatcgtgcgacacaatctaccagaaggctgtgtccgaccg  
 agcttcgcttccacgctgccaagctctgcgacaagatggcgctctttatggtggagggg  
 accaagtccggagcctgctcctcaacatgctgcagaaggacttcacggtgcgcgaggag  
 ctgcagcagcaggacgtggagcgtggctgggcttcatcaccttctgtgtgaggtcttc  
 35      ggcaccatgcgagcagcagcagggcgagcccttccgtgtgctcgtgtgccccatctacacc  
 tgcctcaggagctcttgcaatctcaggatgtgaaggaaagatgctgtcctttgctgctct  
 atggagctgcagagtacaggccggctgctggaggaaacagctgcctgagatgatgacagag  
 ctcttggccagcgacgggacaagatgctgtgccccctcggagtcctatgctgacccggctc  
 40      ctgctccttagaggtcatcgagctccacgctaacagctggaaccctctgacgccccccatc  
 acgagctactacaacagaacctccagaaaactgacagcctgacagccagggggctggca  
 ggcggcccacgggcagctggggccctggtgcacaggggccagatggacaggcgaggaca  
 ggggtggccctggcgggagaaagaaatggggaggaggggcaggcagagtcggtggccagtc  
 45      tggagccagacggggaaggagcaaatccctgagaggagtgcggccgcacaagcccccca  
 gccgagcatgcaagctcacaccaataaagggaagcatgtttcttttcttgggtggccctg  
 gccctcccttctcactcccgcctctcccctccccatcagacccatccccacggagct  
 50      ttgtgtgagggatctcatcgctgtgactcctcggagaccttggcagcctcgacgcgggg  
 gcaccgcttgggtcagaaaggacctcggaaggctgaaaaagtggtcgagagcgggctcg  
 cattgttcccgcagctgtgcagccgcagtcgccaactggcagcaggcgacgtgtagcaga  
 tgtccgggaggacaaggcaggcaggtccccaccagccggccgtaattgacggcctttg  
 55      tcagccatggcagagctgacgctccacctcccacctccaagtcctcctcactgcagcccc  
 cacagcctcaggcctagggggtcaggcgcagcgggggagatggagtttgagttccactt  
 gcactcttttgtttattgtgttttatttttcaaaagtcggttgcctttgaagtctctttgg  
 ccaatgaaaatgcccgtgaggtgatcacacagtcagcactgttgaggacccccggattag  
 60      tgggagatcaaaccagctccccctctagaagaaggattcgagccacagacagcttgccag  
 tagccaattagggtaattggaaaacttctgccccggcggggggtccccgctggaatcctgt  
 gttcctcgccactggcttccagcgcctctgttttctcaaagggtgatactgtcaccact  
 gggaccaagttaaacctggtcctggccccaggggccttgtggcaaacagggcacagaacg  
 agactggcaaatataaaccaaatctagatggtgtcttgcgctccacacgcaggtctta  
 ctggggaaaaggatgggagtggggggctccccaggactcgattttagctaatacgctgtgt  
 cactgccccagctcggacgtagaagcccagccctccgtgagctcttgggaaagggtgaa  
 65      ttactgggtcatggaaggacagtcaggtgaccagcggggtcgccagatgaagcttccc  
 agccgggaaacaagacgggggttcttggcaggccctggctcctggggagcaggccctgttg  
 ttggctggagaggaaggtgtgggggtggaacaggtgtccacatagctccatctctgggggc  
 tggagcacacactttgatgagccccccggaaatgatgtcagagcctagccgcttctta  
 70      tttgctcttttattgagggccgggcaggccctgggtcactttggaggccctcttgggtcca  
 cactggactggccgggaggtgatgggcggggaaggttctcgtgattgattgattctgagt  
 ctgagagtggcgagtggggagaggcttccccagttctctccagcttccctgcagctgca

acctgccctctgggtcccaggtgtggagcctttgcctgtctctaaaaagagcctgttggcg  
 acaaggtgtagggggcacaagtttacctgaaacaggtcagtggtctctcccaagaagcgc  
 acgccacctctgggtccctggccctgaacctgccttcttccctccacaggtttcttcc  
 5 cagactttctcaagtcctcctcactgcccttccctcccagcccagcctgggaacacaga  
 tgccccgcgggttaggagcctcagggaggagcgggctgatgcggggctgctcagggca  
 ggccccagggcgagcttgccatcgtggccaggcagcctccacctgtgcttcagtggcccc  
 tgccccctgaagcatgtgggggttgctcgctaggaggaggcaagggccccgaagagagg  
 agagacctgggagtgaggagctcaggtcagggaggaggcaggggagtggggtctcccagac  
 10 ccaacggtgagctcagagcaagcttcacgcaggacgctccgaaacactgtgtggaggggg  
 ctgtgttgtgggcaccttggggcctgattctccttccctccgaacgggctccttgatggcc  
 tggccacaggggcagctccccattggctgttaggaccagagtgtgaagaagaagtgaat  
 ataaatatgtatacatatataaataatatttttaattacatgtcgtgtcacgggtggctcca  
 gacatactgtttgcctagtttattccactgcttgaaagcgttccctagccaatctgaaca  
 15 acaacactttaagctgtttttctaaatgcaggttgctgctccttttccagatatggaagg  
 aaaacgttaagactatttttttttaagaaacaacagtcaagcctaaaaatttgagaccc  
 cgaggcagcttcccagggagactgctcagacaggaactgcaggacagaagtggatgccc  
 cacagaccctggccccctcccaagtcctccctctctgtggcatgaggaaggccgcgt  
 ccgagttgacctctgaatgtatgtgatgagaggcagagctggatattgcatttctaaggc  
 20 ttgcattgctttccctcgcgggttcttggcgcatggaagaggcgggtccagccatct  
 gatgttgatcctgtctcagtcctccctcactgctgtcaggatgagtttagtcattgttttc  
 tccgaggcggcctgcttgccacagccctgctccccaaggcctgggtggctttgcccgaagct  
 ctgggacgcgagccccagcgaggcccccaacctcaccagacgaggccaggagccccgc  
 25 acctccacgggtagtgccacctcagacccattctctctgttcgtccttccctgaccag  
 tctgtaaacccttactgtttggggatcgtcctgtccatccatgtaaatgtaaatggttggc  
 cgagtcggtattttattctgattgattttttatttctattattttctccgagggatga  
 ggggtggggggtgtgggaagggtaccacagatcaggccggggcagctgtaggggcgggggc  
 ccagacagccaggccgcccaccagagcagccccatggggtgcccagacgcgggcctccaa  
 gaagccaagtcccagctgttttctggcatcagacaccggcccgtgttccctgtcagaca  
 30 gacagactctcaggcctgcttggggagtcgtgtccctcagctgcagggcactgtgttggg  
 aaaccattggctgggcttttgaggacacagatcagaagaaagaaagacaactttcctctg  
 cgcggaacactcacacggaagggtgcccgcctccctgagccggctgggagtgagcagaca  
 ggacctacctcccccagagcaagggtggttcccgccaaagctgcccgggaaccccg  
 ctagtgcgaccacctccctccgtcggtatgtcctgctttccagctgaacccaaactaca  
 35 agtgggtttaaaaaaaataaacaccaccacccaaaaaac  
  
 >ab021179 (TOXMARKER Assignment: 65; SEQ ID NO: 150)  
 gcaggttgggagggaaaagtcgggggaggacgcggaagaggagctgtgggaagggggagga  
 gggagggagggaaaagaggagagacggaggagaaactgagcagagcagagcatcgagccaa  
 40 aggggagatgagtttgtctgtcctctgctgaggctacggccgggcttagggaaactgggag  
 cttgggtggaagcgacaccctggaagtgggaggaggtggcgccgggactttaaccctt  
 gtgggtctgcggcaggggatttaaccctttgtggatctggccctcggaggcagcgtca  
 tggtagttttaacccttccggggtgggtttcagcactggacttacctcatcactct  
 gctcaccaactcctttattgggggtgctccgcttggaggtttgaggccacctccgccc  
 45 tactactgttctgcccgtgcaccccttggacccgctagctggccgactgtgggcgc  
 ttaaccctttactgacttgagctcccagattgcagttggagtttgtgatagaaggact  
 agctaaaggcgtcactgcaggaattacaaactgaagaggactctgttggactgtttttt  
 tttctttttctttttttaagaaaaaccatttttttcccttaaggacttactagccaaaa  
 tttcttaaacttcgaggactctactagccatggccgagccattcttgtcagaatatcaac  
 accagcctcaaactagcaactgtacaggtgctgctgctgtccaggaagagctgaaccctg  
 50 agcggccccagggcgaggagcgggtgcccagggaggacagtaggtggcaatcgagag  
 cgttccccagttgggtggcctgcggggcgagggggaaggagcctggaatcccaac  
 cacctcccttcagacccaggcctgtccagaatctagctgctgagagagggcgagaagg  
 gccagaatggggacgactcgtccgctggcggcgacttcccgccggcgagaagtggaaac  
 cgacgcccagggccgagctgctcgcccagccttgtcatgactccgaggccagtaagttgg  
 55 gggctcctgcccagggggcgaagaggagtggggacagcagcagagacagctggggaaga  
 aaaaacataggagacgcccgtccaagaagaagcggcattggaaaccgtactacaagctga  
 cctgggaagagaagaaaaagttcgacgagaaacagagccttcgagcttcaaggatccgag  
 ccgagatgttcgccaagggccagccggtgcgcgcctataacaccacgcagttcctcatgg  
 60 atgatcacgaccaggaggagccggatctcaaaaccggcctgtactccaagcgggcccgcg  
 ccaaatccgacgacaccagcagatgacgacttcatggaagaagggggtgaggaggatgggg  
 gcagcgatgggagggggagggcagcgagtttctgcagcgggacttctcggagacgt

acgagcgggtaccacacggagagcctgcagaacatgagcaagcaggagctcatcaaggagt  
 acctggaaactggagaagtgcctctcgcgcatggaggacgagaacaaccggctgcggtgg  
 agagcaagcggctgggtggcgacgcgcgtgtgcgaggagctggagctggagctggacc  
 5 ggctgcgcgccgagaacctccagctgctgaccgagaacgaactgcaccggcagcaggagc  
 gagcgccgctttccaagtttgggagactagactgaaacttttttgggggagggggc  
 ggactttttacagtgtatggaatgtaacattatatacatgtgtatataagacagtggacct  
 ttttatgacacataatcagaagagaaatccccctggctttgggtgggtttcgtaaatttag  
 ctatatgtagcttgcgtgctttctcctgttcttttaattatgtgaaactgaagagtgtgt  
 10 tttcttgttttccctttttagaagttttttcccttaatgtgaaagtaatttgaccaagtta  
 taatgcatttttgtttttaacaaatccccctccttaaacggagctataaggtggccaaatc  
 tgagaacaattaaattcatttttagttataataaatttaattttgtaaatgtaacatagt  
 ttcaggtgttcttagagctaatccaaatagtagttgatataattttatgtgactgcatt  
 tttggggaggggtaccgaaatcgtaaattttgtcagtttgcaaaaatatcaatctttaat  
 15 gggagaattttcaatttggcaattttttccttgaatgggtttaagtatgctacaatat  
 agttcaggcaaaaatttaagatgtaattatcttcaatacttaagtgtgcttgccttctagt  
 gccttgggttttcttcttctgatgctggaaaaataaacaaccgggtattgagtgtttaggcg  
 agtggaagtggctacaatccaaaatttttaatttaactctgcctcgccattcaaaagt  
 ctaataacaaaaaatgtaaacctaatgtggcagtttggttaggttagacaactgacagcct  
 20 catttcatttccataagttgggtttcagtaatctcttccctccccccagtaaggctggaa  
 gaggtccttggcaaaccttcttagcgcaagcaatggtagattaatgtgagggcagctct  
 ttaagacgttcagaggttaagaaatactggatttataaagcaaatggctgtttgggggatt  
 ccaaggatttaccataattgtccaattctacgtgctctctataccaaaacaaaaaaagc  
 tatccacctttccatgtgggtcacaactaaaattagaaatgtccctcactgcagatcaaa  
 25 tgtaaagcttccagtttaaggagctaaatgaggtcctcagctgaatgaggaacctgtaca  
 tccccctgacagccctattctaaatcgcttaactatgctgatagctgcttaggttctt  
 gagtagttctgctcttaaacgtaggaggccctgagaactaaattttgccccaaaaataaa  
 aacagaaattatgagattgcctcctgtcattttgggttaaccagtccttcacctgcctg  
 tgtcagtgcttctgagggcaattgcgttgctcaaactactagcacagaggttccctaat  
 30 ttggggccttagaaaccattgtgggccttgggggtccatgaacccccatgaaattatttgta  
 gacttgtagtacatttttctggggagaaggttcaagagattcataagattgtcaaacctc  
 ttgaaaggttcagaacctctgcagggaagggggaagaaaacctcccattaggaagcatg  
 cttttgcagttaaaatggcgatgggtgaggtgatagggacttcaagagtaaaatgcacctt  
 gtattgcataagaagcatacacaatcaataaatacaaggagattataccagtaggactg  
 35 aatcagggccttcaagctggactgagttgggtcctgttctggcacatatgggtccactgga  
 gacaatgtatgattgagcttttcttgggtctaaaaattatattaacatttttttgaaa  
 aaaaaaaaaaaaaaaaaaaaaa

>l41944 (TOXMARKER Assignment: 42; SEQ ID NO: 135)  
 cccgcactaaagacgcttcttccccggcgggtaggaatcccgcggcgagccgaacagttc  
 40 cccgagcgcagcccgcgaccaccacccggccgcacgggcccgttttgtcccccgccgc  
 cgcttctgtccgagagccgcccgcgagggcgcacctcctgaccgcgagcgtcgggtcccaga  
 gccgggcccgggtggggccgaggttagcatctctcgggagccgcaaggcgagagctgca  
 aagtttaattagacacttcagaattttgatcacctaattgttgatttcagatgtaaaagtc  
 45 aagagaagactctaaaaatagcaaatgcttttgagccagaatgccttcattcttcagat  
 cacttaatttggttctcatggtgtatatcagcctcgtgtttgggtatttcatatgattcgc  
 ctgattacacagatgaatcttgcactttcaagatatcattgcgaaatttccgggtccatct  
 tatcatgggaattaaaaaaccactccattgtaccaactcactatacattgctgtatacaa  
 tcatgagtaaacagagaagatttgaaggtgggttaagaactgtgcaaataccacaagatcat  
 50 tttgtgacctcacagatgagtggagaagcacacacgagggcctatgtcacctgcctagaag  
 gattcagcgggaacacaacgttgttcagttgtctcacacaatttctggctggccatagaca  
 tgtcttttgaaccaccagagtttgagattgttgggttttaccacacattaatgtgatgg  
 tgaattttccatctattgttgaggaagaattacagtttgatttatctctcgtcattgaag  
 aacagtcagaggggaattgttaagaagcataaacccgaaataaaaggaaacatgagtggaa  
 55 atttcacctatatcattgacaagttaattccaaacacgaactactgtgtatctgtttatt  
 tagagcacagtgtgagcaagcagtaataaagtctcccttaaaatgcaccctccttccac  
 ctggccagggaatcagaatcagcagaatctgccaaaataggaggaataattactgtgtttt  
 tgatagcattgggtcttgacaagcaccatagtgacactgaaatggattgggttatatatgct  
 taagaaatagcctccccaaagtcttgaggcaaggtctcgctaagggtggaatgcagtgg  
 60 ctattcacaggtgcagtcataatgcactacagctctgaaactcctgagctcaaacagtcgt  
 cctgcctaagcttccccagtaggtgggattacaagcgtgcacccctgtgccccagtgatt  
 aagttttattatgtagaaaaataagagcaaacagtcacagctgatatggactctctctctc

ttttttttttttttaagaattttcataacttttttagcctggccatttcctaacctgcc  
 ccgttgggaagccatggatgatggtggaggtcatttacatcaacagaaagaagaagtgtgg  
 gattataattatgatgatgaaagtgatagcgaactgaggcagcgccaggacaagtggc  
 5 ggtggctataccatgcatggactgactgtcaggcctctgggtcaggcctctgccacctct  
 acagaatcccagttgatagacccggagtcgaggaggagcctgacctgacctgaggttgat  
 gtggagctccccacgatgccaaaggacagccctcagcagttggaactcttgagtggggccc  
 tgtgagaggagaaagagttccactccaggaccctttccgaagaggactacagctccacg  
 gaggggtctggggggcagaattaccttcaatgtggacttaaactctgtgtttttgagagtt  
 10 ctgatgacgaggacagtgacgacttagaagccctctgatgctatcgtctcatctggaa  
 gagatggttgacccagaggatcctgataatgtgcaatcaaaccatttgctggccagcggg  
 gaagggacacagccaacctttccagccctcttcagagggcctgtgggtccgaagatgct  
 ccatctgatcaaagtgacacttctgagtcagatgttgacctggggatggttatataatg  
 agatgactccaaaactattgaatgaacttggacagacaagcacctacagggttctttgtc  
 15 tctgcatcctaacttgctgccttatcgtctgcaagtgttctccaaggggaaggaggaa  
 actgtgggtgttccctttctccaggtgacatcacctatgcacattcccagtatggggacca  
 tagtatcattcagtgacattgtttacatattcaaagtgggtgcactttgaaggaagcacatg  
 tgcacctttctttacactaatgcacttaggatgtttctgcatcatgtctaccaggagc  
 aggggtccccacagtttcagaggtgggtccaggaccctatgatattctctctctctctc  
 20 ttttttttttttttttgagacagagtcctcgtctgtcgcaccaagctggagcgcaatggg  
 tgatcttggtcactgcaacatccgcctcccggttcagggtgattctcctgacctcagcct  
 cctcgcaagtagctgggattacaggcgccctgccaccatgcctagcaaatttttgtatatt  
 ttagtggagacaggattttaccatgttggccaggctgggtctcgaactcctgacctcaagt  
 gatctgcctcctcagcctcgtaaagtgtgggttacaggggtgagccgctgtgctgg  
 25 ctggccctgtgatatttctgtgaaataaattgggccaaggtgggagcagggaaagaaaag  
 gaaaatagtagcaagagctgcaagcaggcaggaaggaggaggagagccaggtgagcag  
 tggagagaagggggccctgcacaaggaaacagggaagagccatcgaaagtttcagtcggt  
 gagccttgggcacctcaccatgtcacatcctgtctcctgcaattggaattccacctgt  
 ccagccctcccagttaaagtggggaagacagactttaggatcacgtgtgtgactaatac  
 30 agaaaggaaacatggcgctcggggagagggataaaacctgaatgccatattttaagttaa  
 aaaaaaaaa

>bc015897 (TOXMARKER Assignment: 93; SEQ ID NO: 159)  
 gccttctaagcctctgaatgcaattacatgtatttcagaacattctaagaagtaataa  
 atcatcatccagatgtacaaacaaaagatgataagctcaaaaactcagttttggcccaag  
 35 gtccctggtgctaccagttcagctgcaaatacctgtaaggtacagccacttactcttaaag  
 agactgctgaaagttttggaagccaccaaagaagaatgggaaatgaacacatcagtg  
 tccacctgaaaactcagactgtatccaagcagatgttaactctgatgattacaaggggtg  
 ataaagtataccatccagaaacaggaaggaaaaacgagaaagagaaagttggaagggaagg  
 gcaagcatctgttgactgttgatcagaaacgtggagagcatgttgtctgtggcagcacac  
 40 gtaataatgagtcagagagcactttggatttagaaggcttccaaaatcccacagctaaag  
 agtgtgagggacttgccactttagataacaaggctgatcttgatggagaaagtaacgaag  
 gtactgaggaactagaagactctctaaccactttaccactcagtaacagggccagacat  
 cagaatgattccctctgatgaagaggaggagacgacgaagaagaggaggaggaagaag  
 45 aacctaggctcaccattaaccaaagggaagatgaagatggcatggctaataagatgagt  
 tagacaacacctacactgggtcaggggatgaggacgcctatctgaagaggatgatgagt  
 taggcgaagctgctaagtatgaagacgtgaaagaatgtggaaaacatgtagaagagctc  
 tcctagtggaaacttaataaaaataagtcctcaaggaaagaaatgtatgtgaagaaaaaatt  
 cacctgtggatcagtcgtatttttttatgaattcagtaaaacttatcttcaccaaaggca  
 50 agtctcctacggtagtgtgcagcttatgcaaacgagagggtcatctaagaaggactgtc  
 ctgaagacttcaaaagaatccagctagaacctctgccaccattaacacccaagtttttaa  
 atatcttagatcaagctctgtatccagtggtataaggatttttctccaacaattatagaag  
 atcaggctcgtgaacatattcggaacacctagaagtttcataagacaggactttccag  
 gaactaaattgagcctgtttggctcctccaaaatggatttgggttcaaacagagtgacc  
 55 ttgacgtctgtatgacaattaatggacttgaaactgctgagggattggactgtgtcagaa  
 ctattgaagaattagcaagagtcctcagaaaacattcaggctctgagaaacatcttaccta  
 ttacaacagcaaaggtgccaattgtgaagttcttccatttgagaagtggtctggaagtag  
 atatcagtttgtataacacattggcccttcataacacaaggcttttatctgcttattccg  
 ccattgatcccagagtgaagtatttgtctataccatgaaagtatttacaagaatgtgtg  
 60 atattggatgcacatagaggcagcttatcatcgtatgcataactcttatggtgctat  
 atttctccagcagaggaatccaccagtcattcctgtccttcaagagatatcaaagggtg  
 aaaagaaacctgaaatatttgttgatggctggaatatttatttttttgatcaaatagatg

aactgcctacctattgggtcagaatgtggaaaaatacagaatctgttgggcagttatgggt  
 tgggccttcttcggtttctacacagaggaatttgattttaagaacatgttattagcatca  
 ggagaaaaagtctgcttacaacttttaagaacagtggaacctcaaaatacattgttattg  
 5 aagatccctttgatttgaatcataatcttggagctggattatcaaggaaaatgacaaatt  
 ttataatgaaggcttttatcaatggtagaagagtatttggatttctgtcaagggtttc  
 caaaggactaccctcaaaaatggaatacttttttgatccagatgtgttaactgaaggag  
 agctggcccaaatgatagatgttgtcgaatttgtggaaaaatcggacacttcatgaagg  
 actgtcctatgaggagaaaagtaagacggcgagatcaggaagatgccctgaaccaa  
 10 gataccctgagaacaaggaaaaaagaagcaagaggacaaagaaattcacaacaagtaca  
 cagaaagggaggtgtcaacaaaagaagataagccatacagtgcacacctcagaaagcca  
 agccaatgcgggcagctgctgacctggggagggagaagatcctcaggccaccagtagaaa  
 aatggaagagacaggatgacaaagacttaagagaaaaacgttgttttatttgtggaagag  
 aagggcacattaaaaaggaatgcccacagtttaaaggctcttcaggtagcctttccagta  
 aatatatgactcagggaagacctcagcgaagaggaccagcaggaatcatgaggggaagg  
 15 aaaaatgcagcactctaaatggccactcaggcgttcttattcactcggaaaattaggttca  
 tttcacaggacacagcagtgtagatcaggcttcaacttaacatttaagggaatgtcaga  
 tttttttttaatttaataatgaaattgttaatgaggaaaaatttttaatatagtcttatctac  
 cacacatccccatagatttaaggatttttaataagaaagtcagatgtatgtatttaagcca  
 cgtaaaagaaaaaatataactatggaccggtattcagtgaatacagtttcatgggtttt  
 20 aattctttcaagcacattaaaaatgggtgtgctgataaaacccaagtaaatcaacccctt  
 ttccgtataaatccattttttgttttgaagaggggaaattatatttattgttgttactg  
 aatcctggtgtgaaagcatatcagatatgtatgaactgctactgctgtacttccgattta  
 cggacatcattttattgtctattttagacgtgataacatgaacatgagtacctatttatg  
 tgggccttcagtggtgggcagtgccactcagggtctctggggtttccctctctaatttta  
 25 agtaaattgacataataactactatgcttataaaaaatgaaagtaaggaaaaacaagtagtct  
 gtttgccactaaaaacattttcaaggaaaaataaaaatgaaagtactttttactttttat  
 gatactcagaaattagatgaagaacttttaaaatgtctgaagatcaaagaggttatctc  
 tgccagtcacaagtgtggctgggtgtcattctgggtctgactggagccctcctggactgtt  
 tctttaatttcaaaagccctgcagacatagtacctggtcagaactatgcctcgggtttatt  
 30 tatcattttgaaataaaatcaaaatttcaacctgtaaaaaaaaaaaaaaaaaaaaaaaaaa  
 aa  
 aaaaaa

>bc002457 (TOXMARKER Assignment: 95; SEQ ID NO: 160)  
 ggcacgaggccgccttctgcatcgcggttcggcggttccacctagacacctaacagtc  
 35 gcggagccggccgcgtcgtaggggggtcggcacggggagtcgggcggtcttgtgcatctt  
 ggctacctgtgggtcgaagatgtcggacatcggagactggttcaggagcatcccggcgat  
 cacgcgctattgggttcgccgccaccgtcgccgtgcccttggtcggcaaacctcggcctcat  
 cagcccgccctacctcttctctggcccgaagccttcccttatcgctttcagatttggag  
 40 gccaatcactgccaccttttatttccctgtgggtccaggaactggatttctttatttgggt  
 caatttatattctttatcagttatctacgcgacttgaaacaggagccttttgatgggag  
 gccagcagacttttattcatgctcctctttaaactggatttgcacgtgattactggcct  
 agcaatggatatgcagttgctgatgattcctctgatcatgtcagtttattgtctgggc  
 ccagctgaacagagacatgattgtatcattttgggttggaaacagatttaaggcctgcta  
 45 ttaccctgggttatccttggattcaactatatcatcggaggctcggtaatcaatgagct  
 tattggaaatctgggttgacatctttatttttctaatgttcagataaccaatggactt  
 gggaggaagaaattttctatccacacctcagtttttgtaccgctggctgccagtaggag  
 aggaggagtatcaggatttgggtgtgccccctgctagcatgaggcgagctgctgatcagaa  
 tggcggaggcgggagacacaactggggccagggtttcgacttggagaccagtgaagggg  
 50 cggcctcgggcagccgctcctctcaagccacatttccctccagtgctgggtgcgcttaac  
 aactgcgttcttggttaacactgttggacctgaccacactgaatgtagtcttccagtacg  
 agacaaagtttcttaaatccgaagaaaaatataagtgttccacaagtttcacgattctc  
 attcaagtccttactgctgtgaagaacaaataaccaactgtgcaaatgcaaaactgaaaa  
 aa  
 55 aaa

>af161504 (TOXMARKER Assignment: 97; SEQ ID NO: 154)  
 ggtgccttatgcggtgatttttaatgataggtgtcatatataggacggagtaactctgttta  
 60 cattctgttcttctcgatgcactcacaagcgggtaactaggtagacaagaaaacaaagatc  
 ttattcaaaaagggtcttacagcaacccaacgtctcatcttcccatagtaaagatgacgg  
 cgccttgaggtaagctacaggcaacaccacttccgcgtttctcttgcgccttggtccaag



atggcggatgaagccacgcgacgtgttgtgtctgagatcccgggtgctgaagactaacgcc  
 ggaccccgagatcgtgagttgtgggtgcagcgactgaaggaggaatatcagtcacctatc  
 cggatgtggagaacaacaagaatgctgacaacgattggttccgactggagtccaacaag  
 5 gaaggaaactcgggtggtttggaaaatgctggtatatccatgacctcctgaaatatgagttt  
 gacatcgagtttgacattcctatcacatatcctactactgccccagaaattgcagttcct  
 gagctggatggaaagacagcaaagatgtacaggggtggcaaaatatgcctgacggatcat  
 ttcaaacctttgtgggcccaggaatgtgccccaaatttgactagctcatctcatggctctg  
 gggctgggtccatggctggcagtggaatccctgatctgattcagaagggcgtcatccac  
 10 caciaagagaaatgcaaccaatgaagaatcaagccactgaggcagggcagagggacctt  
 gataggctacgatactattttccctgtgcatcacacttaactcatctaactgcttccccgg  
 acaccctccacctctagttgttactaagtagctgcagtaggcattgctggggaagaaaca  
 aacacacaccaaacagtactgctacttagtttctaaggctgcacaggggaagggaaagact  
 gggctttggacaatctagaggtaatatccgccccaggtggagcaacatgcgattc  
 15 tggaggcacggggtaactgaaagtgagtacatatagcttttctggtttctggagataac  
 ccatcaataaaaagctgcttccctctggtaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

>af085361 (TOXMARKER Assignment: 60; SEQ ID NO: 146)  
 ggtcgcttgggtggctccgtctgttgcctcgcccgcggtgccatcatggcggacgcg  
 20 gccagtcaggtgctcctgggtccggtctcaccatcctgtcccagccgctcatgtacgtg  
 aaagtgctcatccaggtgggatagagcctcttccccaacaataggacgaaatattttt  
 gggcggcaagtgtgtcagcttccctggtctcttttagttatgctcagcacattgccagtatc  
 gatgggaggcgcggttgttcacaggcttaactccaagactgtgttcgggagtccttggga  
 actgtggtccatggttaaagttttacagcattaccaggagagtgaaggggtgaggagtta  
 25 ggacctggaaatgtacagaaagaagtctcatcttcccttgaccacgttatcaaggagaca  
 actcgagagatgatcgctcgttctgtcgtcctaccctcatcacacatcccttccatgtgatc  
 actctgagatctatggtacagttcattggcagagaatccaagtactgtggactttgtgat  
 tccataataaccatctatcggaagagggcattctaggatttttgcgggtcttgttccct  
 cgccttctaggtgacatccttcttctgtggctgtgtaactcactggcctacctcgtaac  
 30 acctatgcactggacagtggtgttctaccatgaatgaaatgaagagttattctcaagct  
 gtcacaggattttttgcgagtatgttgacctatcccttgtgcttgtctccaatcctatg  
 gctgtcaacaactgtggtcctgtggtggatgccctccttactccccaatatatacgtct  
 tggatagactgttgggtgatgctacaaaaagaggggaatatgagccgaggaaatagctta  
 ttttccggaaggcccccttgggaagacttattgttgtgacctgaaaaatgttaattga  
 35 agatgtggggcagggacagtgacatttctgtagtccagatgcacagaattatgggagag  
 aatgttgatttctatacagtggtggcgcttttttaataatcatttaactcttgggaaat  
 taaaaaaaaaaaaaaaaaaaaaaaaaaaaa

>af110646 (TOXMARKER Assignment: 54; SEQ ID NO: 144)  
 ccaagatggcggcggcagacacagcagcagccagttatcgggaaaggcagacagtggt  
 40 ctttgaagcgtatgttgaatttcaatgtgcctcatattaaaaacagcacaggagaaccag  
 tatggaaggtaactcatttatgacagatttggccaagatataatctctcctctgctatctg  
 tgaaggagctaagagacatgggaatcactctgcatctgcttttacactctgatcgagatc  
 ctattccagatgttccctgcagtatactttgtaatgccaactgaagaaaatattgacagaa  
 45 tttgccaggtcttgcgaatcaactatatgaatcatattattttaaattttatttctgcta  
 tttcaagaagttaaactggaagatattgcaaatgcagcgttagcagctagtgcagtaaacac  
 aagtagccaagggtttttgaccaatatctcaattttattactttggaagatgatattgttg  
 tattatgtaataaaaaaaggagcttgtttcatatcgtgccattaacaggccagatatca  
 50 cagacacggaaatggaaactgttatggacactatagttgacagcctcttctgcttttatg  
 gtactcggggtgatgttccataatcagatgttcaagaggaacagcagcagaaatggtag  
 cagtgaactagacaagaaacttcgagaaaaatctaagagatgcaagaaacagtcctttta  
 cagggtgatacacttggagctggccaattcagcttccagaggcccttattagtccttgttg  
 acagaaacatagatttggcaactcctttacatcatacttggacatatcaagcattggtgc  
 55 acgatgtactgtgatttccatttaaacagggttaatttggagaagatcttcaggagtgga  
 actctccagctgggtgctagaccacaaagagaaaaaacaagaagtcttatgatttaactccg  
 ttgataaattttggcaaaaacataaagggaagtcattcccagaagttgcagaatcagttc  
 agcaagaactagaatcttacagagcacaggaagatgaggtcaaacgacttaaaagcatta  
 tgggactagaaggggaagatgaaggagccataagtatgctttctgacaataccgctaagc  
 60 taacatcagctgttagttcttggcagaactccttgagaaaaaagacttattgatctcc  
 atacaaatgttgccactgctgttttagaacatataaaggcaagaaaattggatgtatatt  
 ttgaatatgaagaaaaaataatgagcaaaactactctggataaatctcttctagatataa  
 tatcagaccctgatgcaggaactccagaagataacatgaggttgtttcttattattata

5 taagcacacagcaagcaccttctgaggctgatttggagcaatataaaaaagctttaactg  
atgcaggaaaccttaatcctttacaatatatcaaacagtgaaggcttttaccagatgg  
cctcagctccggccagctatggcagcactaccactaaaccaatgggtcttttatcacgag  
tcatgaatacaggatcacagtttgtgatggaaggagtgaagaacctgggttttgaaacagc  
10 aaaatctacctgttactcgtattttggacaatcttatggagatgaagtcaaaccgccgaaa  
ctgatgactatagatatatttgatcccaaaatgctgcggggcaatgacagctcagttccca  
gaaataaaaatccattccaagaggccattgtttttgtgggtgggaggaggcaactacattg  
aatatcagaatcttgttgactacataaaaggggaaacaaggcaaacacattttatatggct  
gcagtgagctttttaatgctacacagttcataaaacagttgtcacaaacttggacaaaagt  
aacacagaagaaccttactatgataatctacttggaaatgtggataaatgtaaaaagaaga  
aaagttagaagagcaatatgtttccttctctgtaacagtgctcctaacagtgaaaaatcaga  
gttatttgttaatttttaaggaaattatatacttaatatgtattgattaaaagaacatt  
tcagaaataaaatttcaacattgaaaaaaaaaaaaaaaaaaaa

15 >bc015110 (TOXMARKER Assignment: 66; SEQ ID NO: 151)  
ggggagctctgcgaggggcccggagcgcggcggagccatgcagtaccgcaccccgggccg  
gcgggcgggcgccgtgggggtgccgctgtacgcgcccacgcgctgctgcaaccgcacac  
ccgacgcccccttttacatcgaggacatcctgggcccggggcccgccgcccacgcccgc  
20 cccacgctgccgtcccccaactcctccttcaccagcctcgtgtccccctaccggaccccg  
gtgtacgagcccacgccgatccatccagccttctcgcaccactccgcccgcgctggcc  
gctgcctacggaccccgcggttctcgggggccctctgtaccccttcccgcggacggtgaac  
gactacacgcacgccttgcctcgccacgacccctgggcaaacctctactctggagcccc  
ttcttgagaggcctctgcataaaaggaaaggcgccaggtgagattctccaacgaccag  
accatcgagctggagaagaaattcgagacgcagaaatatctctcgcggccgagaggaa  
25 cgtctggccaagatgctgcagctcagcgagagacaggtcaaaacctggtttcagaatcga  
cgcgctaaatggaggagactaaaacaggagaacctcaaagcaataaaaaagaagaactg  
gaaagtttggacagttcctgtgatcagaggcaagatttgccagtgaaacagaataaaggt  
gcttctttggatagctctcaatgttcgccctccccctgctcccaggaagaccttgatca  
gagatttcagaggattctgatcaggaagtggacattgagggcgataaaagctattttaat  
30 gctggatgatgaccactggcattggcatgttcagaaaactggatttaggaataatgtttt  
gctacagaaaatcttcatagaagaactggaaggctatataagaaagggaatcaattctct  
ggtattctggaacctaataaatatttgggtgcactgctcaattaacaaacctacatggaga  
ccttaattttgacttaacaaatagtttatgtactgctcttaggttgttttgataaagtga  
cattatagtgtattaaattcttcccccttttaaaaaaacagttagtgttttactatttat  
35 aaaaaattaattttgaactttttgttaaatttttaagttatagctttaaagggttttaata  
ggaccttcttgaacgacttttctgtaatctgtttatctccacttaatggaaaggcaaaag  
gggtaccccaaatccagaggtgcctacatttcaggcagccttgaggtattttaaaaggaa  
aacattctttacttttatatgacattcttatactgctgtctcaaatccaaaaacatttca  
gagctcttgcctcagagatgtgtgttcttttgcagagatatggttgatgagaatctta  
40 aatgcttgttttgactatcacttagtacctgttgaccaagggtgttaaggggatagtac  
ctcccaattcaagcagagaaaactgacctgactaaagttaatcgagatgaactagaagtc  
acaggttaataaatgtaagtagattgtagatactgttttatatcaacaatgtttataa  
tgttatatagaattgttactgtaaaaaaatggccaaatgtgttttttttttaataa  
gtaacttgactataaaataaaagccgtccgtgggacgactgacctcgtaaaaaaaaaaaaa  
45 aa  
aaaaaaaaaaaaaaaaaaaaaaaaaaaa

>nm\_004046 (TOXMARKER Assignment: 39; SEQ ID NO: 132)  
gtcccagtcagtcgggaggctgcggctgcagaagtaccgctgcggagtaactgcaaagat  
50 gctgtccgtgcgcgttgcgtgcggccgtggtccgcgcccttcccgccgggcccggactggt  
ctccagaaaatgctttgggttcatctttcattgctgcaaggaaacttccatgcctctaacac  
tcatcttcaaaagactgggactgctgagatgtcctctattcttgaagagcgtattcttgg  
agctgataacctctgttgatcttgaagaaactggcgctgtcttaagtattgggtgaggtat  
55 tggccgcgtacatgggctgaggaatgttcaagcagaagaaatggtagagtttcttcagg  
cttaaagggtatgtccttgaacttgaacctgacaatgttgggtgttgcgtgtttggaaa  
tgataaactaattaagggaaggagatatagtgaaggagacaggagccattgtggacgttcc  
agttgggtgaggagctgttgggtcgtgtagtgtatgcccttggtaatgctattgatggaaa  
gggtccaattgggtccaagacgcgtaggcgagttgggtctgaaagcccccggtatcattcc  
60 tcgaatttcagtgcggaaccaatgcagactggcattaaggctgtggatagcttgggtgcc  
aattgggtcgtggtcagcgtgaactgattattggtgaccgacagactgggaaaacctcaat  
tgctattgacacaatcattaaccagaaacgtttcaatgatggatctgatgaaaagaagaa

gctgtactgtatTTtattgttgctatttggtcaaaagagatccactgttgcccagttggtgaa  
gagacttacagatgcagatgccatgaagtacaccattgtgggtgctgggtacggcctcgga  
tgctgccccacttcagtagcttggtccttactctggctgttccatgggagagtatttttag  
5 agacaatggcaaacatgctttgatcatctatgacgacttatccaaacaggctgttgctta  
ccgtcagatgtctctgttgctccgcccagccccctggctcgtgaggcctatcctgggtgatgt  
gttctacctacactcccgggttgctggagagagcagccaaaatgaacgatgcttttggtgg  
tggtccttgactgctttgccagtcatagaaacacaggctgggtgatgtgtctgcttacat  
tccaaacaaatgtcatttccatcactgacggacagatcttcttgaaacagaattgttcta  
10 caaaggatccgcccgtcaattaacgttggtctgtctgtatctcgtgctggatccgctgc  
ccaaaccagggtatgaagcaggtagcaggtaccatgaagctggaattgggtcagtatcg  
tgaggttgctgcttttgcccagttcggttctgacctcgtgctgccactcaacaactttt  
gagtcgtggcgtgcttactgagttgctgaagcaaggacagtatctcccatggctat  
tgaagaacaagtggctgttatctatgcgggtgtaaggggatcttgataaactggagcc  
15 cagcaagattacaaagtttgagaatgctttctgtctcatgtcgtcagccagcaccagc  
cttgttgggcactatcagggtgatggaaagatctcagaacaatcagatgcaaagctgaa  
agagattgtaacaaatttcttggtgatttgaagcttaaaactcctgtggattcacatca  
aataccagttcagttttgtcattgttctagtaaattagttccatttgtaaaagggttact  
ctcatactccttatgtacagaaatcacatgaaaaataaaggttccataatgcaaaaaaaa  
20 aaaaaaaaaaaaaaaaaaaaaa

>s52379 (TOXMARKER Assignment: 58; SEQ ID NO: 147)  
atggatcccaactgctcctgcgcgcgggtgactcctgcacctgcgcggctcctgcaaa  
tgcaaaagagtgc aaatgcacctcctgcaagaaaagctgctgctcctgctgccctgtgggc  
25 tgtgccaaagtgtgccagggtgcacatctgcaaggggcgtcggacaagtgcagctgctgc  
gctga

>k03193 (TOXMARKER Assignment: 40; SEQ ID NO: 136)  
gccccggcgccgcccgcgcgcagaccggacgacaggccacctcgtcggcgctccgcccag  
30 tccccgcctcgccgccaacgccacaaccaccgcgcacggccccctgactccgtccagtat  
tgatcgggagagccggagcagctcttcggggagcagcagatgcgacctccgggacggcc  
ggggcagcgtcctggcgtgctgggtgcgtctgcccggcgagtcgggtcttgaggaa  
aagaaagtttgcaaggcagagtaacaagctcagcagttgggcacttttgaagatcat  
ttctcagcctccagaggatgttcaataactgtgaggtggtccttgggaaatttggaatt  
35 acctatgtgcagaggaattatgatcttctccttcttaaagaccatccaggaggtggctggt  
tatgtcctcattgcccctcaacacagtgaggcgaattcctttggaaaacctgcagatcatc  
agaggaaatatgtactacgaaaattcctatgccttagcagctctatctaactatgatgca  
aataaaaccggactgaaggagctgcccatgagaaatttacaggaaatcctgcatggcgcc  
gtgcggttcagcaacaacctgccctgtgcaacgtggagagcatccagtgggcgggacata  
40 gtcagcagtgactttctcagcaacatgtcagtgacttccagaaccacctgggcagctgc  
caaaagtgtgatccaagctgtcccaatgggagctgctgggtgagaggagagaaactgc  
cagaaactgaccaaatacatctgtgccagcagtgctccggcgctgccgtggcaagtc  
ccagctgactgtgccacaaccagtgctgcaggctgcacaggcccccgaggagcagac  
tgcttggtctgccgcaaatccgagacgaagccacgtgcaaggacacctgccccccactc  
45 atgctctacaacccccaccagtagcagatggatgtgaacccgagggcaatacagcttt  
gggtccacctgcgtgaagaagtgtcccgtaatatgtggtgacagatcacggctcgtgc  
gtccgagcctgtggggccgacagctatgagatggaggaagacggcgctccgcaagtgtaa  
aagtgcgaagggccttgccgcaagtggtgaacggaataggattgggtgaatttaaagac  
tactctccataaatgtacgaatattaaacacttcaaaaactgcacctccatcagtggc  
gatctccacatcctgcccgtggcatttaggggtgactccttcacacatactcctcctctg  
50 gatccacaggaactggatattctgaaaaccgtaaaggaaatcacagggtttttgctgatt  
caggcttggcctgaaaacaggacggacctccatgccttgagaacctagaaatcatacgc  
ggcaggaccaagcaacatggtcagttttctcttgagtcgtcagcctgaacataacatcc  
ttgggattacgctccctcaaggagataagtgtgagatgtgataatttcaggaaacaaa  
aatttgtgctatgcaaatataaactggaaaaaactgtttgggacctccggtcagaaa  
55 accaaaattataagcaacagaggtgaaaacagctgcaaggccacaggccaggtctgccat  
gccttgtgctcccccgagggtgctggggcccgagcccagggactgcgtctcttgccgg  
aatgtcagccgagggcaggggaatgcgtggacaagtgaaccttctggaggggtgagccaagg  
gagtttgtggagaactctgagtgcatcagtgccaccagagtgctgcctcaggccatg  
aacatcacctgacagggggaccagacaactgtatccagtggtgccactacattgac  
60 ggccccactgcgtcaagactgcccggcaggagtcagggagaaaacaacacctgggtc  
tggaagtacgcagacgcccggcatgtgtgccacctgtgccatccaaactgcacctacgga

5 tgcactggggccaggtcttgaaggctgtccaacgaatggaagctacatagtgctcacttt  
ccaagatcattctacaagatgtcagtgactgaaacatgcagggcggtgttgagtggtga  
aggaatcttgacaagtgttttgaagatagcattttgctaagtccttgaggtcactggtcc  
tcaaagcggcatggcgcatggcggtgggtctgtgccacatgccagctgtgtgacctctg  
10 agactccacttcttccgtgctgaaaataaagaaggagttttactaaggacaaacaagat  
aatgaatgtgaaactgctccatgaaccccaagaattatgcacatagatgcatcattaa  
gatgcaagccatcgagttaccacctggcatgcttaactgtaaagagtgggtcaaagta  
aactgaattggaaaatccaaagttatgcagaaaaacaataaaggagatagtaaaaagggt  
taacgagccagtccaggggaagcgaagaagacaaaaagagtccttttctgggccaagttt  
gataaattaggcctcccgaccctttgctctgttgctttatcaactctactcggcaataac  
aat

15 >bc017832 (TOXMARKER Assignment: 41; SEQ ID NO: 137)  
gccccctccctccgccccgccccgccccgccccgctcagtcctggcaggcaggcaggcaatcg  
gtccgagtggtgtgctggtcttccagctctcccgctcggcgctcttccctcctcctcccggt  
cagcgctcggcggtgacccggcgggcgcgagtccttgcgggaggggcgacaagagctga  
gcgggcgcccgccgagcgtcgagctcagcgcgggcgaggcgggcgggccggcagccaac  
atggcgggcgggcgggcgggcgggcgggcgggcgggcgagatggtccgcgggcagggtgttcgac  
20 gtggggcgcgctacaccaacctctcgtagatcgggcgaggcgccctacggcatggtgtgc  
tctgcttatgataatgtcaacaaagttcgagtagctatcaagaaaatcagccccctttgag  
caccagacctactgccagagaacctgagggagataaaaaatcttactgcgcttcagacat  
gagaacatcattggaatcaatgacattattcgagcaccaaccatcgagcaaatgaaagat  
gtatatatagtagcagacctcatggaaacagatctttacaagctcttgaagacacaacac  
ctcagcaatgaccatatctgctattttctctaccagatcctcagagggttaaaatatatc  
25 cattcagctaacgttctgcaccgtgacctcaagccttccaacctgctgctcaacaccacc  
tgtgatctcaagatctgtgactttggcctggcccggtgttgagatccagaccatgatcac  
acagggttcttgacagaatatgtggccacacgttggtacagggtccagaaattatgttg  
aattccaagggtacaccaagtcattgatattttggtctgtaggctgcattctggcgaaa  
atgctttctaacaggcccatctttccagggaagcattatcttgaccagctgaaccacatt  
30 ttgggtattcttggtatccccatcacaagaagacctgaattgtataataaaatttaaagct  
aggaactatttgctttctctccacacaaaaataaggtgccatggaacaggctgttccca  
aatgctgactccaaagctctggacttattggacaaaatgttgacattcaacccacacaag  
aggattgaagtagaacaggctctggccacccatatctggagcagtagtacgacccgagtg  
35 gacgagcccatcgccgaagcaccattcaagttcgacatggaattggatgacttgccctaa  
gaaaagctcaagaactaatttttgaagagactgctagattccagccaggatacagatct  
taaatttgtcaggtacctggagttaataacagttagctctagcaaggaggcgctgcctt  
ttgcttctagaatattatgttccctcaaggtccattattttgtattcttttccaagctcct  
tattggaaggtatttttttaaattagaattaaaaattatttagaaaaaaataaaaaaa  
40 aaaaaaaaaaaaaa

>ah002741 (TOXMARKER Assignment: 98; SEQ ID NO: 161)

45 >aw293499 (TOXMARKER Assignment: 51; SEQ ID NO: 142)  
tttttttttttttttttgcgggtcgagagactcccgcttgccaagggtttttattgtggtcc  
cgcggggcaggaggtatgcatggcatagcgaagcagagagccggaggcagccatcgggcac  
ctagaacggtgcagagttggcccaggagcgtggcggggcaggcgccctgcacctgccttg  
ctcgcccagcagaccctccgggctccagcctggcggggcccagcgtccaccttggtgggc  
50 ccagggtcagatcttgcccagggtggagtgggcgctcggcctgctcctcttggtgggggtc  
cggaaactgcctccccagggggcttggtgggcacggggggcagcctctggtccctccgg  
agcaggtacagggccagcaggtgggcagggggccagcagccccagcaccaggcccgagg  
cccaggatggggggaaccgcacgggccccggggacctccacgggcccgggtg

55 >aj300837 (TOXMARKER Assignment: 46; SEQ ID NO: 138)  
ggcacgagggtggcctctgcggctaggccggctcgagactcccgggcgccgaggcgctgc  
cgcccgctcgccgccccacgcgaaggaccacgcgccccgcccggccagcctctcagcg  
ctcccatgatcgcccggtgccttttggtgtggaagcctccgcagagttggtggttcca  
ggattttactcagaatgacgttaggaagagaagtgtgtctcctcttcaggcaatgtctt  
cctatactgtggctggcagaaatgttttaagatgggatctttcaccagagcaaattaaaa  
60 caagaactgaggagctcattgtgcagaccaaacagggtgtacgatgctgttggaatgctcg  
gtattgaggaagtaacttacgagaactgtctgcaggcactggcagatgtagaagtaagtg

5 ataatagtggaaggaccatgctagactttccccagcatgtatcctctgacaaagaagtac  
 gagcagcaagtagagaagcagacaaaagactttctcggtttgatattgagatgagcatga  
 gaggagatatatttgagagaattgttcatttacaggaaacctgtgatctggggaagataa  
 aacctgaggccagacgatacttggaaggtcaattaaaaatggggaaaagaatgggctcc  
 10 atcttctgaaacaagtacagaatgaaatcaatgaagaaaagaatgagtgaagctat  
 gtattgattttaacaaaaacctcaatgaggatgataccttctgtattttccaaggctg  
 aacttggtgctcttctctgatgatttcattgacagtttagaaaagacagatgatgacaagt  
 ataaaattaccttaaaatatccacactatttccctgtcatgaagaaatgttgatccctg  
 aaaccagaagaaggatggaaatggcttttaatacaaggtgcaaagaggaaaacaccataa  
 15 ttttgacagcagctactcccactgccaacaaaggtggccaaactactcggttatagcacac  
 atgctgacttcgtccttgaaatgaacactgcaaagagcacaagccgcgtaacagccttcc  
 tagatgatttaagccagaagttaaaaccttgggtgaagcagaacgagagtttattttga  
 atttgaagaaaaaggaatgcaaagacaggggttttgaatatgatgggaaaatcaatgcct  
 gggatctatattactacatgactcagacagaggaaactcaagtattccatagaccaagagt  
 20 tcttcaaggaatacttcccaattgaggtggctcactgaaggcttgctgaacacctaccagg  
 agttgttgggactttcatttgaacaaatgacagatgctcatgtttggaacaagagtgtta  
 cactttatactgtgaaggataaagctacaggagaagtattgggacagttctatttggacc  
 tctatccaagggaaggaaaatacaatcatgcggcctgcttcgggtctccagcctggctgcc  
 ttctgcctgatggaagccgatgatggcagtggtgacctcggtggaacttctcacagc  
 25 cagtgccaggtcgctccctctctcctgagacacgacgaggtgaggacttactttcatgagt  
 ttggtcacgtgatgcacagatttgtgcacagactgattttgcacgatttagcggaacaa  
 atgtggaactgactttgtagaggtgccatcgcaaatgcttgaagaaatgggtgtgggacg  
 tcgattccctccgaagattgtcaaaacattataaagatggaagccctattgcagacgatc  
 tgcttgaagaaactgttgcttcttaggtggctcaacacaggtcttctgacctgcgccaga  
 30 ttgttttgagcaagttgatcagctctcttcataccaacacatcgctggatgctgcaagt  
 aatatgccaaatactgctcagaaatattaggagttgcagctactccaggcacaaatatgc  
 cagctacctttggacatttggcagggggatacagatggccaatattatggatatctttgga  
 gtgaagtattttccatggatatgttttacagctgttttaaaaaagaagggataatgaatc  
 cagaggttggaatgaaatacagaaacctaatcctgaaacctgggggatctctggacggca  
 35 tggacatgctccacaatttcttgaaacgtgagccaaacccaaaagcgttcctaataagta  
 gaggcctgcatgctccgtgaactggggatctttggtagccgtccatgctctggaggacaag  
 tcgacatcaccatgtgttactggcctggaaactgaagggagttttgcaagtgaagaaatga  
 gatttctattgacatcctttgttttctaatttttaaaattataaagatgtaaatggaat  
 40 tataaatactgtgacctaagaaaagaccactagaaaagtaattgtactataaaatttcat  
 aaaactggatttggatttctttttatgaaagtttcatatgaatgtaacttgattttttact  
 attataatctagataaatatgatataagagggctaagaatttttaaatgaaatcatatata  
 tgatataatttgatccttctgtatcttgaagtttgtacttgggatttctggactgata  
 aatgaatcatcacattcttctggtaaatattttcttgagctctgtgtcaactttgatcc  
 45 tttgtctcccaggaaggtgtgacctctccttgcctgcatacctcaaggccaggggaata  
 tgccctcagtgatgcatttatctttgtatatcaggccgcgcatgattcccaactttctgccac  
 acttaaatcagttcctccatttccagttttgtcttttctgtctaaagttcagtcaaagag  
 tatcaaaaaa  
 >aj242973 (TOXMARKER Assignment: 49; SEQ ID NO: 140)  
 50 gcgccgcgctcgacgtgacagccggtacgcccgggtttgggcaacctcgattacgggcgg  
 cctccaggcccgccagcagcgcggccgcggccgcggcccgcccgcccgcccgcccggtt  
 ccggccgcggaccctcctctgcggtccggctgcggctccgctgcccggtagcgccgctc  
 cccgggaccacccttcggctggcgccctcccatgctctcgccaccggagggttgcc  
 agtcctcctcctccacagcctcttcccggtcccagagatgggcaactcggcctcgaaca  
 55 tcgtcagccccagggagccttgccgggcccgaaggaaacagaccctgtagcgcccaaac  
 atcatgtcaatggcaacagacagtcgaacctttcccagagggaacacagatggctgtat  
 ttggaatgggatgtttctggggagctgaaaggaaattctgggtcttgaaaggagtgtatt  
 caactaaagtgtggttttgcaggaggtatacttcaaatcctacttataaagaagtctgct  
 cagaaaaaactggccatgcagaagctcgccgagtggtgtaccagccagaacacatgagtt  
 60 ttgaggaactgctcaaggtcttctgggagaatcacgacccgacccaaggatgcgccagg  
 ggaacgaccatggcactcagtagcctcgccatctacccgacctctgccaagcaaatgg  
 aggagccctgagctccaaagagaactacaaaagggttcttccagagcacggcttcggcc  
 ccatcactaccgacatccgggagggacagacttctactatgcggaagactaccaccagc  
 agtacctgagcaagaaccccaatggctactgcggccttggggggcaccggcggtgctgcc  
 cagtggttattaaaaataattgctccccacatggtgggcctttgaggttccagtaaaaa  
 tgctttcaacaaattgggcaatgcttgtgtgattcacaatcgaggcatttaagtgacaca

5 aagtacaaaggaattttatacagattgggtttaccgaagtataatctataggaggcgcgat  
 ggcaagttgataaaaatgtgacttatctcctaataagttatgggtgggagtgaggctgtgcg  
 gtttccctgtgtcttctgggtctgagtgaagatagcagggatgctgtgttcaccccttctt  
 ggtagaagctaaggtgtgagctgggaggttgctggacaggatgggggaccccagaagtcc  
 tttatctgtgtctctgcccgcagtgcttacaatttgcaaactgtatagcctcagtg  
 actcattcgtgaaatccttcgctttacca

10 >aj002030.1 (TOXMARKER Assignment: 126; SEQ ID NO: 166)  
 ccggtgatggcggtggtgatggggacgtgaagctaggcaccctggggagtggcagcgag  
 agcagcaacgacggcgagcagagagtcaggcgacgcgggagcggcagcggaaggggga  
 ggctgggcgggcgggcggttggcgcttctgacggggggcggggaaatgctgctgaacgtg  
 gcgctggtggctctggtgctgctgggggacctaccggctgtgggtgcgctgggggcgggcg  
 ggtctggggggcgggggcgggggcgaggagagccccgccacctctctgcctcgcatg  
 15 aagaagcgggacttcagcttggagcagctgcgcagtagcagcggtcccgaaccgcgc  
 atcctgctcgcggtcaatgggaaagtcttcgacgtgaccaaaggcagcaagttctacggc  
 ccggcggtccatatggaatatttgcgtgtagggatgcctccagaggactggccacattt  
 tgcctagataaagatgcacttagagatgaatatgatgatctctcagatttgaatgcagta  
 caaatggagagtggttcgagaatgggaaatgcagtttaaagaaaaatagattatgtaggc  
 20 agactcctaaaaccaggagaagaaccatcagaatatacagatgaagaagataccaaggat  
 cacaataaacaggattgaactttgtaacaaccaaagtcaggggccttcagaactgcaat  
 tcttactccctttcacagactgtccggagtctttgggtttgattcacctgctgcgaaaaa  
 cattcaacaattgtgtacaagataaattaatctcactatgaagatttgaataactagac  
 attatttatgctgcgaactcatttgggtgagttgtttgtaattgtctagtgggcttcat  
 25 catcctgaaaagaaggagacagggatttttttaagagcaagaaagtcacaataattactt  
 ctttcccttcttttttcttcttcttcttcttcttcttcttcttcttcttctttaaata  
 ttgaagacaaccagatatgtatttgcactcaagtgtacagatctcctcaagaaacatca  
 agggactcctgtgtcacatactgtgtttttattttaacatgggtgagggaggcgacctga  
 tcaggggaggtgggggtacacatcaatttgagttgttcaggctactgaaacattaaaatg  
 30 tgaattcccaaacttttctttttggctttgtcagggaaaagaaaaatatctttataaaga  
 aatctttggaaattaggagaaggaatttcagggtgggtttaagtcagagctagttccccc  
 cagaagatcatttgaaaccagtttttatcccttctcttcccttcccttccctaaatca  
 aatcaatattaattgtgccttatttcacttaacatagacttgaattatttttagggaaag  
 cccctataatgaattcagaaatcactacaagcagcattaagactgaagttggaatattct  
 35 gttgaccataaaaacttgatatcattctgtgtatataagaatgtaaaaggaatattacagt  
 gtttaactgccatatatgtaataacacaaactcaattagcattgtaatggccaaatgc  
 tcccccatgcttttctgttttcaaaaaaattgaaaaacaaatcaactcttatccccaa  
 gctgcctaatttttaggagctgtgacctccacatctcactggtgtgggtgcaggggctgt  
 ggagtggtgtcagtatggatgtgtctgaatgtgtgagggccttggaggggactcttctg  
 40 cagatactgtaataacaagtaccattttaataaagcatgtacaataaaccaaaaaaaaaa  
 aaaaaaaaaaaaaa

45 >ak056165.1 (TOXMARKER Assignment: 110; SEQ ID NO: 167)  
 ggactcgagcgctccgattggagttagggcctgcttgtctgcgtgctgcgaagtcgcg  
 ctgccccggggccctagtcgttgggttccagggtccttcacgttccattcccaggctgg  
 tctgagctccggggccgtggtcccgctgcctcctccggctcgtcgtgcggaagctgcgacg  
 caggcagaccatggcagagttctccagaaacgggggaagcggcgtagcgacgaagggtc  
 gggcagcatggtggacttccctcctggccaatgcccgcctggtgctgggcgtgggcggggc  
 50 tgctgtgctgggcattgccaccctggccgtgaagcgggttcattgacagggcactagccc  
 gcgggatgaggatgacaccaaggcagacagctggaaggaaactgagcctgctcaaggccac  
 accacacctgcagccccggcctccacctgctgcccttagccagccagtggtgacctggc  
 cccctcgtcgtctgccccagaaggcctgcagaaactgatcctgaggtgacaccacagct  
 cagctccccagcaccgctgtgtctgacactgcaggagaggtgctggccttcgagcggga  
 ccgtgtgaccatcccagcagcccaggtggcttggccaaacagctggctggcgacatcgc  
 55 cctggagctgcaggcctactttcgagcaagttcccggaactgcccttggggcattcgt  
 gcctggggggccgctctacgacgggctgcaggcgggggctgcggaccatgtgcgtctcct  
 ggtgccactggtgctggagccgggctgtggagcctggtgccgggctggacactgtggc  
 gagggaccctcgctgctgggcccgtgcgcaggacgcagcttgagttctgccccgtgggag  
 cagcccctgggaccgcttccgtggtcgggggctacttctcctcccgcgtcctgctggagct  
 60 actccgcaaggcgtggtgcttctgtcaactggccggccattggcagccttctcggggtg  
 cctgatccggccagcatggcctcggaggagctgctgctcgaggtgcagcacgaacgcct  
 ggagctcactgtggctgtgcttgtggcagtcctgggggtcgatgctgacgaccgctcct

cttggcctggccctggagggctggcggggaacctctggtcgaggacctgtatccagt  
ggaggtctgtaggctgcgagccctggacgacctgacgtctgggactcgccggcggtgct  
gctgctgctgtgtgctgtctgccgtggttgctcggctctggggcagctaggccggggtca  
cctgaccaggtggtcctgctgtctgggggaggacaacgtggattggacggaggaggcctt  
gggtgagcgcttctgcaagccctggagctgctcatcggcagcctggagcaggccagcct  
gccctgccacttcaacccagcgtgaacctcttcagcagcttgctgtaggaggagattga  
cgacattggctatgcgctatacagtggcctacaggagcccaggggctgctctaggtggg  
tggaacgggtggttgccattgtttctaatgtctggggagctgacccacctccctccag  
ggatttgaatagtggttttctctagcttttggcagacaagagggtacattactta  
aaccagggtcaggtatgtgcttgggctatggtggccataaacctgagccagagagc  
ttgggtcactgtcacctgagtgcagctgggctgcctcaggcagcttggagtccagccat  
tcttgaagcaccgtttcagctcttggggccaacccaggaccttggctctgtccatca  
ccagcaaccaatccaccaacagaatgtggtttctgccatcctgggcagaagctgaaggcc  
agcttcacatttctgctgagagaaggtgacttaacgcctttccggccctagctccaggc  
gttttgaggcgctctggtgcctgatggtaggtatggtgtgtttgttctgtccccagggg  
tgaggtcacctggtgccccctgaaggacagatttttggtgtttaaaggatggcattttcct  
gctgtcttctgtgcgtttagttttcttctgctgagcgggagctcagatgacttggcacc  
cctgataacctcagggcaggcccttttccctccagccaggtgagtgtttcttcaggca  
gctgaggggtcctgggggagctgaggctctgtgctgcacccccagccacagctggggcat  
ctcactggagctgttccaggccccactggagagcagaggacctgatccccactagagag  
gtccggtgtgcacagccggcctcccagtggtgccaataaactgactctcagctgatggct  
gtattctgactttgaagcctgttaagaggtagcaaggggctagaggaggagattccac  
ctccccctccaagtgacctcctcctgctctggtatccttccctttgaaacgaagctca  
gcttcgaagatgtgaacaagaataaaaggaaaaatttctaattgatatat

>nm\_005968 (TOXMARKER Assignment: 114; SEQ ID NO: 162)  
cgcgcgggaacagccagtcggtgcctaacgcgagtgatctcgagagagaagcgatcaa  
cagctgcgggtctgcgcctgcgcggcgacggggcgtggccgcgggcgagtggggccaagg  
aggcagccgggagcgggggcgaggtgttactggttgctcgggtcacgtgggcgcgcag  
cagaccgcggtgcagccggttcgctcacacaaagcccagacgcggagaaaatggcggcag  
gggtcgaagcggcggcggaggtggcggcgacggagatcaaaatggaggaagagagcggcg  
cgcccggtgcccagcgggaacggggctccgggccctaagggtgaaggagaacgacctg  
ctcagaatgagaagaggaaggagaaaaacataaaaagaggaggcaatcgctttgagccat  
atgccaatccaaataaaaagatacagagccttcattacaacataccttttgatgtgaaat  
ggcagtcacttaagacctggttaaaagaaaagtgtggtgagtgaaacatacgtggagctct  
taatggacgctgaaggaaagtcaaggggatgtgctgttgttgaattcaagatggaagaga  
gcatgaaaaaagctgcggaagtcttaacaagcatagcttgagcggaagaccactgaaag  
tcaaagaagatcctgatggtgaacatgccaggagagcaatgcaaaaggatggtggtacga  
ctggtgggatgggtatgggaccaggtggccagggaatgattactatcccaccagtatcc  
taaataatcccaacatcccaaatgagattatccatgcattacaggctggaagacttgga  
gcacagtattttagcaaatctggattataaagttggctggaagaaactgaaggaaagtat  
ttagtatggctggtgtggtggtccgagcagacattcttgaagataaaagtggaaaaagtc  
gtggaataggcatggttacttttgaacagtcattgaagctgtgcaagctatatctatgt  
tcaattggccagctgctatttgatagaccaatgcagctcaagatggatgagagggccttac  
caaaaggagatttcttccctcctgagcgtccacaacaacttccccatggccttggtggta  
ttggcatggggttaggaccaggagggcaacccattgatgccaatcacctgaataaaggca  
tcggaatgggaacataggtcccgcagggaatgggaatggaaggcataggatttggaaata  
ataaaatgggagggaatggaggggccccttgggtggtggtatggaaaaacatgggtcgattt  
gatctgggatgaacatgggcaggataaatgaaatcctaagtaatgactgaagagaggag  
agatcattgcaaaagcaggaggagggtggaggtggaggaagcgtccctgggatcgagagga  
tggtcctggtgattgaccgctcgggggtgcccggcatggagcgcatgggcgcgggcttg  
gccacggcatggatcgctgggtccgagatcgagcgcatgggctggctgatcgaccgca  
tggtgctccgtggagcgcattgggtcctcggcattgagcgcagtgggcccgctgggctcgacc  
acatggcctccagcattgagcgcagtgggccagaccatggagcgcattggctctggcgtgg  
agcgcagtggtgcccgcagtggtctcgcccttgagcgcagtgggcgtcccatcgaccgtg  
tgggccagaccattgagcgcagtggtcttgccgtggagcgcagtgggcctgccatcgagc  
gcatgggctgagcagtgagcgcagtggtgcccgcaggtatgggagctggcctggagcgcga  
tgggccccctgatggatcgcatggccaccggcctggagcgcagtgggcgccaacaactctgg  
agcggatgggctggagcgcagtgggcgccaacagcctcgagcgcagtgggcctggagcgcga  
tggttgccaacagcctcgagcgcagtgggccccgccatgggccccgctgggctggcga  
ttgagcgcagtgggcctggccatgggtggcggtggcggttgcagctttgaccgtgcatcc

```
>nm 005968 (TOXMARKER Assignment: 114; SEQ ID NO: 162)
```

5 agatggagcgtggcaacttcggaggaagcttcgcaggttcctttggtggagctggaggcc  
 atgctcctggggtggccaggaaggcctgccagatatttgtgagaaatctgccattcgatt  
 tcacatggaagatgctaaaggacaaattcaacgagtgccggccacgtgctgtacgccgaca  
 tcaagatggagaatgggaagtccaaggggtgtggtgtggttaagttcgagtcgccagagg  
 10 tggccgagagagcctgccggatgcatgaagctgagtgccgagagattgacg  
 ttcgaattgatagaaacgcttaagcagttgccttttttaaacatcgatacgagacctcg  
 aatttgtatttttcttgttaaccattttaatttgttggtggatgtataaagatgttta  
 aaaaattcagttgctttttggggtaatttgaattacttttttaatgactgggggttcatt  
 tgactgtttgcattgagattgcaatgtgcgcaattttttttagttagtggtgcatcttgtt  
 15 gacatcgaatatgactttgataataaataccggttcctgaaaaaaaaaaaaaaaaaaaa  
 aaa

>ak026121 (TOXMARKER Assignment: 44; SEQ ID NO: 138)  
 15 agttgctgtggaggccctggcagcggtgcagcaggggtgtgagcgccaccgttgcccacct  
 tctggacctggcaggcagcgcgggtgcgactgggagctggcgtagcccctctgagccaca  
 ggagccgctgggtgcaggacctgcaggctgctgtggccgcccgtccagagtgcctgcacga  
 gctgttgaggtttgcccgagcgcgggtgggcaatgctgccacacatctgacctgacct  
 gcatgccaagcttagccggcagctgcagaagatggaggacgtgcaccagacgctggtggc  
 20 acatggtcagggcctcgacgctggccgggaggctctggagccaccttgaggacctgga  
 ccggctggtggcctgctcgcggtgtgcccaggagcgaagcagctggcctccttctc  
 gcacggcaatgcctcactgctcttcagacggaccaaggccactgccccggggcctgaggg  
 ggggtggcaccctgcacccaacccactgacaagaccagcagcatccagtacagaccct  
 gccctcacccttaagttcacctcccaggactcgccagatgggcagtagcagaacagcga  
 25 ggggggctggatggaggactatgactacgtccacctacaggggaaggaggagtttgaaa  
 gaccagaaggagctgctggaaaaggcagcatcacgcggcagggaagagccagctgga  
 gttgcagcagctgaagcagtttgaaacgactggaacaggaggtgtcacggcccatagacca  
 cgacctggccaactggacgccagcccaacccctggccccggggcgaaacaggcggcctggg  
 gccctcgagccggcagctgctgctcttctacctggagcagtgtagggccaacctgaccac  
 30 actgaccaacgcctggacgccttctttaccgcccgtggccaccaaccagccgccaagat  
 ctttgtggcgacagcaagttcgtcatcctcagcgcccacaagctggtgttcatcgggga  
 cacactgtcacggcaggccaaggctgctgacgtgcgcagccaggtgacctactacagcaa  
 cctgctgtgcgacctcctgcgggcacgtggccaccaccaaggccgctgccttgagta  
 cccatcgcttcccgggccagacatggtggagagggtcaaggagctgggccaacagcac  
 35 ccagcagttccgcgctccttagccagctggcagccgctgaggggtggtgacccagga  
 gggaggcaggggaggggtgcggcggtcccagctccctggctcccatgtcaagagtcgctg  
 tgccacaggcttagggacaggaccccagctctgctcggtcctggtgccctggatgcca  
 ggaatctgtatataattatggccgggcagggtgtggggccatgcctcctcaggagccgaa  
 gccagggggccggccagtgcccttcccagcatgcaccacggggccgggttggtcacca  
 40 gacgggggtgagtgtagggctcctgcagcctgcaggacctcgtgccaccccgagggtg  
 agcctggtcccacgaggggtgcctgtgtcccctgacagggccagtgagtttggtgtgtcct  
 ccgctttccaggagaagaacctgaagaactatttttcgttatttggtttccaatcattt  
 gactaagagctctccatttaataaaagtttttaaaaggaaaaaaaaaaaaaaaaaaaa

>u90545.1 (TOXMARKER Assignment: 119; SEQ ID NO: 152)  
 45 acgctccgcccacgcgtccgcccacgcgtccggtcggggccagagcgcaggtgtacctg  
 gcggcctgctggagcacctgaccgcccagatcctggagctggctggcaacccggcccg  
 gacaagaagaccgcatcatcctgcgccacctgtagctggccattcgcaacggcgaggag  
 50 cttacaagctgctggggaagtaccatcgcgaggcggtgtcctgcccacattcag  
 ggctgcttctgccccagaagaccaagagccaccacaaggccaagggtgaaaaccattca  
 ctaggagaggagaaacacaatggccaccaagacagagttgagtcaccacagcaaggagag  
 caagaacgcacaagatatgcaagtggatgagacactgatccccaggaaagggtccaagttt  
 atgttctgctcgctatggaatagccctcgtcttacatttctgcaatttcacaacgatagc  
 55 acaaaatgtcatcatgaacatcaccatggttagccatggtcaacagcacaagccctcaatc  
 ccagctcaatgattcctctgaggtgctgctgtgactcatttggtggcctaagttaaacg  
 cccaaagagtcttctgcaaaagtctcaatacttgggggtcagtttgcaatttgggaaaa  
 gtggggccctccacaagaacgaagcagactctgcagcattgctttatcaggaatgttact  
 gggatgctttactgccatcctcataggtggcttcattagtgaaccccttgggtggccctt  
 60 tgtcttctatcttggagggtgttggtgtgtctgctgccttctctgggttggtgtgat  
 ttatgatgaccccttttctatccatggataagcacctcagaaaaagaatacatcatatc  
 ctcttgaaacaacaggtcgggtcttctaagcagcctcttcccatcaaagctatgctcag  
 atctctacccatttggtccatagttaggtgtttcagccatcaatggttagttagcac



```

aatggttgatatacataccaacttacatcagctctgtgtaccatgttaacatcagagacaa
tggacttctatctgcccttccttttattgttgctgggtcataggcatgggagggcta
tctggcagatttcttcttaacaaaaagtttagactcatcactgtgaggaaaattgccac
aatttttaggaagtctccctcttcagcactcattgtgtctctgccttacctcaattccgg
ctatatcacagcaactgccttgctgacgctctcttgcggaattaagcacattgtgtcagtc
agggatttatatcaatgtcttagatattgtccaaggattccagttttctcatgggagc
atcaagaggattttcgagcatagcacctgtcattgtaccactgtcagcggatttcttct
tagtcaggaccctgagtttgggtggaggaatgtcttcttcttgctgtttgccgttaacct
gttaggactactcttctacctcatatttgagaagcagatgtccaagaatgggctaaaga
gagaaaactcactcgtttatgaagttatccaccttgatggaaaagtcattaggcaccg
tattgcataaaatagaaggcttccgtgatgaaaataccagtgaaaagatttttttttctt
gtggctcttttcaattatgagatcagttcattattttattcagacttttttttgagagaa
atgtaagatgaataaaaaattcaataaaaatgataactaagaaaaaaaaaaaaaaaaa

```

# **Example 10: Identification of a TOXMARKER 76, 135, 147, 151, 152, 154, and 162 for Accurate Prediction of Hepatotoxicity**

In order to determine the minimal number of markers required for prediction, backward selection from a larger set of putative markers was used. Each set of smaller markers was evaluated on cross-validation by linear regression and the smallest set that was significantly accurate ( $p < 0.001$ , Fisher's exact test) was selected. The TOXMARKER genes that make up the TOXMARKER 76, 135, 147, 151, 152, 154, and 162 is listed in Table 6 below

**Table 6. TOXMARKER 76, 135, 147, 151, 152, 154, and 162**

PTS Code	Rat ID No.	TOXMARKE R	SEQ ID NO	Gene Name	Accuracy	P. Value
pts2.3014511.1	scr_gb-bi294409_1	42	42	IFNAR-2	0.771552	4.12E-13
pts2.3013420.1	scr_gb-af069306_1	59	59	Transaldo lase		
pts2.3015170.1	scr_gb-bi288503_1	65	65	Clp-1		
pts2.3011880.1	scr_gb-d86383_2	66	66	Hex		
pts2.3015871.2	scr_gb-bm986259_1	97	97	Novel		
pts2.3012511.2	scr_gb-aa899865_3	76	76	Novel		
pts2.3017180.2	cszr_204152648_1915 21095	71	71	Novel		

“P.value” is the probability level that the observed classification is random.

“Accuracy” is the number of correct predictions divided by the number of samples (total number of predictions). It is a proportion of how often the disclosed TOXMARKER 76, 135, 147, 151, 152, 154, and 162s are accurate in screening for toxicity.

## References *in vivo*:

1. Kedderis, G.L., *Biochemical basis of hepatocellular injury*. Toxicol Pathol, 1996. 24(1): p. 77-83.
- 5 2. Seeman, P., *The membrane actions of anesthetics and tranquilizers*. Pharmacol Rev, 1972. 24(4): p. 583-655.
3. Rosser, B.G. and G.J. Gores, *Liver cell necrosis: cellular mechanisms and clinical implications*. Gastroenterology, 1995. 108(1): p. 252-75.
- 10 4. Neuman, M.G., *Apoptosis in diseases of the liver*. Crit Rev Clin Lab Sci, 2001. 38(2): p. 109-66.
5. Kaplowitz, N., *Cell death at the millennium. Implications for liver diseases*. Clin Liver Dis, 2000. 4(1): p. 1-23, v.
6. Oinonen, T. and K.O. Lindros, *Zonation of hepatic cytochrome P-450 expression and regulation*. Biochem J, 1998. 329(Pt 1): p. 17-35.
- 15 7. Zimmerman, H.J., *Hepatotoxicity: The adverse effects of drugs and other chemicals on the liver*. Second Edition ed. 1999.
8. Horn, K.D., et al., *Biomarkers of liver regeneration allow early prediction of hepatic recovery after acute necrosis*. Am J Clin Pathol, 1999. 112(3): p. 351-7.
- 20 9. Cotran RS, K.V., Robbins SL, *Robbins Pathological Basis of Disease*, W.B. Saunders, Editor. 1994. p. 833.
10. Achord, J.L., *Cirrhosis of the liver: new concepts*. Compr Ther, 1989. 15(2): p. 11-6.
11. Dragan, Y.P., et al., *Implications of apoptosis for toxicity, carcinogenicity, and risk assessment: fumonisin B(1) as an example*. Toxicol Sci, 2001. 61(1): p. 6-17.
- 25 12. Newberne, P.M., et al., *The role of necrosis in hepatocellular proliferation and liver tumors*. Arch Toxicol Suppl, 1987. 10: p. 54-67.
13. Troyanskaya, O., et al., *Missing value estimation methods for DNA microarrays*. Bioinformatics, 2001. 17(6): p. 520-5.
14. Breksa, A.P., 3rd and T.A. Garrow, *Recombinant human liver betaine-homocysteine S-methyltransferase: identification of three cysteine residues critical for zinc binding*. Biochemistry, 1999. 38(42): p. 13991-8.
- 30 15. Avila, M.A., et al., *Reduced mRNA abundance of the main enzymes involved in methionine metabolism in human liver cirrhosis and hepatocellular carcinoma*. J Hepatol, 2000. 33(6): p. 907-14.
- 35 16. Schepers, L., et al., *Presence of three acyl-CoA oxidases in rat liver peroxisomes. An inducible fatty acyl-CoA oxidase, a noninducible fatty acyl-CoA oxidase, and a noninducible trihydroxycoprostanoyl-CoA oxidase*. J Biol Chem, 1990. 265(9): p. 5242-6.

17. Wilkinson, D.G., *Eph receptors and ephrins: regulators of guidance and assembly*. Int Rev Cytol, 2000. 196: p. 177-244.
18. Stein, E., et al., *Nck recruitment to Eph receptor, EphB1/ELK, couples ligand activation to c-Jun kinase*. J Biol Chem, 1998. 273(3): p. 1303-8.
- 5 19. Stewart, M.J. and G. Thomas, *Mitogenesis and protein synthesis: a role for ribosomal protein S6 phosphorylation?* Bioessays, 1994. 16(11): p. 809-15.
20. Sturgill, T.W. and J. Wu, *Recent progress in characterization of protein kinase cascades for phosphorylation of ribosomal protein S6*. Biochim Biophys Acta, 1991. 1092(3): p. 350-7.
- 10 21. Vanmuylder, N., et al., *Heat shock protein HSP86 expression during mouse embryo development, especially in the germ-line*. Anat Embryol (Berl), 2002. 205(4): p. 301-6.
22. Dale, E.C., et al., *Murine 86-kDa heat shock protein gene and promoter*. Cell Stress Chaperones, 1997. 2(2): p. 87-93.
- 15 23. Kanamura, S. and J. Watanabe, *Cell biology of cytochrome P-450 in the liver*. Int Rev Cytol, 2000. 198: p. 109-52.
24. Trottier, E., et al., *Identification of CYP2B14P and CYP2B16P, two apparent pseudogenes in the rat cytochrome P450 2B (CYP2B) subfamily*. Biochem Pharmacol, 1996. 52(6): p. 963-5.
- 20 25. Schaller, M., et al., *Cloning and expression of succinic semialdehyde reductase from human brain. Identity with aflatoxin B1 aldehyde reductase*. Eur J Biochem, 1999. 265(3): p. 1056-60.
26. Weenink, S.M. and A.M. Gautam, *Antigen presentation by MHC class II molecules*. Immunol Cell Biol, 1997. 75(1): p. 69-81.
- 25 27. Martinez, O. and B. Goud, *Rab proteins*. Biochim Biophys Acta, 1998. 1404(1-2): p. 101-12.
28. Waisman, D.M., *Annexin II tetramer: structure and function*. Mol Cell Biochem, 1995. 149-150: p. 301-22.
29. Paulusma, C.C., et al., *Congenital jaundice in rats with a mutation in a multidrug resistance- associated protein gene*. Science, 1996. 271(5252): p. 1126-8.
- 30 30. Suzuki, H. and Y. Sugiyama, *Excretion of GSSG and glutathione conjugates mediated by MRP1 and cMOAT/MRP2*. Semin Liver Dis, 1998. 18(4): p. 359-76.
31. Balduyck, M., et al., *Human leucocyte elastase (HLE) preferentially cleaves the heavy chain H2 of inter-alpha-trypsin inhibitor (ITI)*. Biol Chem Hoppe Seyler, 1993. 374(9): p. 895-901.
- 35 32. Diarra-Mehrpour, M., et al., *Human inter-alpha-trypsin inhibitor: full-length cDNA sequence of the heavy chain H1*. Biochim Biophys Acta, 1992. 1132(1): p. 114-8.

33. Sun, J., et al., *A new family of 10 murine ovalbumin serpins includes two homologs of proteinase inhibitor 8 and two homologs of the granzyme B inhibitor (proteinase inhibitor 9)*. J Biol Chem, 1997. 272(24): p. 15434-41.
34. Safadi, F.F., et al., *Cloning and characterization of osteoactivin, a novel cDNA expressed in osteoblasts*. J Cell Biochem, 2001. 84(1): p. 12-26.
35. Oldak, M. and J. Malejczyk, *[Signal transduction mechanisms induced by epidermal growth factor receptor (EGFR) and their role in apoptosis regulation]*. Postepy Hig Med Dosw, 1999. 53(2): p. 315-29.
36. Hayes, J.D. and L.I. McLellan, *Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress*. Free Radic Res, 1999. 31(4): p. 273-300.
37. Kaas GEN, J.M., Orrenius S., *Cyclosporine A protects hepatocytes against prooxidant-induced killing*. Biochem Pharmacol, 1992. 44: p. 1995-2003.
38. Schulte-Frohlinde D, S.C., *Radiolysis of DNA and model systems in the presence of oxygen*, in Oxidative Stress. 1985: Orlando. p. 11-40.
39. Tribble, D.L., T.Y. Aw, and D.P. Jones, *The pathophysiological significance of lipid peroxidation in oxidative cell injury*. Hepatology, 1987. 7(2): p. 377-86.

#### References *in vitro*:

1. Kedderis, G.L., *Biochemical basis of hepatocellular injury*. Toxicol Pathol, 1996. 24(1): p. 77-83.
2. Seeman, P., *The membrane actions of anesthetics and tranquilizers*. Pharmacol Rev, 1972. 24(4): p. 583-655.
3. Rosser, B.G. and G.J. Gores, *Liver cell necrosis: cellular mechanisms and clinical implications*. Gastroenterology, 1995. 108(1): p. 252-75.
4. Neuman, M.G., *Apoptosis in diseases of the liver*. Crit Rev Clin Lab Sci, 2001. 38(2): p. 109-66.
5. Kaplowitz, N., *Cell death at the millennium. Implications for liver diseases*. Clin Liver Dis, 2000. 4(1): p. 1-23, v.
6. Oinonen, T. and K.O. Lindros, *Zonation of hepatic cytochrome P-450 expression and regulation*. Biochem J, 1998. 329(Pt 1): p. 17-35.
7. Zimmerman, H.J., *Hepatotoxicity: The adverse effects of drugs and other chemicals on the liver*. Second Edition ed. 1999.
8. Horn, K.D., et al., *Biomarkers of liver regeneration allow early prediction of hepatic recovery after acute necrosis*. Am J Clin Pathol, 1999. 112(3): p. 351-7.
9. Cotran RS, K.V., Robbins SL, *Robbins Pathological Basis of Disease*, W.B. Saunders, Editor. 1994. p. 833.

10. Achord, J.L., *Cirrhosis of the liver: new concepts*. Compr Ther, 1989. 15(2): p. 11-6.
11. Ying, T.S., D.S. Sarma, and E. Farber, *Role of acute hepatic necrosis in the induction of early steps in liver carcinogenesis by diethylnitrosamine*. Cancer Res, 1981. 41(6): p. 2096-102.
- 5 12. Dragan, Y.P., et al., *Implications of apoptosis for toxicity, carcinogenicity, and risk assessment: fumonisin B(1) as an example*. Toxicol Sci, 2001. 61(1): p. 6-17.
13. Newberne, P.M., et al., *The role of necrosis in hepatocellular proliferation and liver tumors*. Arch Toxicol Suppl, 1987. 10: p. 54-67.
- 10 14. Troyanskaya, O., et al., *Missing value estimation methods for DNA microarrays*. Bioinformatics, 2001. 17(6): p. 520-5.
15. Hastie, T., R. Tibshirani, and J.H. Friedman, *The elements of statistical learning : data mining, inference, and prediction*. Springer series in statistics. 2001, New York: Springer. xvi, 533.
16. Venables, W.N. and B.D. Ripley, *Modern applied statistics with S-PLUS*. 3rd ed. Statistics and computing. 1999, New York: Springer. xi, 501.
17. Shambaugh, G.E., 3rd, *Urea biosynthesis I. The urea cycle and relationships to the citric acid cycle*. Am J Clin Nutr, 1977. 30(12): p. 2083-7.
18. Hatae, N., Y. Sugimoto, and A. Ichikawa, *Prostaglandin Receptors: Advances in the Study of EP3 Receptor Signaling*. J Biochem (Tokyo), 2002. 131(6): p. 781-4.
- 20 19. Pancholi, V., *Multifunctional alpha-enolase: its role in diseases*. Cell Mol Life Sci, 2001. 58(7): p. 902-20.
20. Norflus, F., S. Yamanaka, and R.L. Proia, *Promoters for the human beta-hexosaminidase genes, HEXA and HEXB*. DNA Cell Biol, 1996. 15(2): p. 89-97.
- 25 21. Vinogradov, A.D., *Mitochondrial ATP synthase: fifteen years later*. Biochemistry (Mosc), 1999. 64(11): p. 1219-29.
22. Oldak, M. and J. Malejczyk, *[Signal transduction mechanisms induced by epidermal growth factor receptor (EGFR) and their role in apoptosis regulation]*. Postepy Hig Med Dosw, 1999. 53(2): p. 315-29.
- 30 23. Peyssonnaud, C. and A. Eychene, *The Raf/MEK/ERK pathway: new concepts of activation*. Biol Cell, 2001. 93(1-2): p. 53-62.
24. Hardy, M.P., et al., *The soluble murine type I interferon receptor Ifnar-2 is present in serum, is independently regulated, and has both agonistic and antagonistic properties*. Blood, 2001. 97(2): p. 473-82.
- 35 25. Prejean, C. and O.R. Colamonici, *Role of the cytoplasmic domains of the type I interferon receptor subunits in signaling*. Semin Cancer Biol, 2000. 10(2): p. 83-92.
26. Nelms, K., et al., *The IL-4 receptor: signaling mechanisms and biologic functions*. Annu Rev Immunol, 1999. 17: p. 701-38.

27. Ruest, P.J., et al., *Mechanisms of CAS substrate domain tyrosine phosphorylation by FAK and Src*. Mol Cell Biol, 2001. 21(22): p. 7641-52.
28. Yoshizumi, M., et al., *Src and Cas mediate JNK activation but not ERK1/2 and p38 kinases by reactive oxygen species*. J Biol Chem, 2000. 275(16): p. 11706-12.
- 5 29. Sakane, F. and H. Kanoh, *Molecules in focus: diacylglycerol kinase*. Int J Biochem Cell Biol, 1997. 29(10): p. 1139-43.
30. Swannie, H.C. and S.B. Kaye, *Protein kinase C inhibitors*. Curr Oncol Rep, 2002. 4(1): p. 37-46.
- 10 31. Walikonis, R.S., et al., *Densin-180 forms a ternary complex with the (alpha)-subunit of Ca<sup>2+</sup>/calmodulin-dependent protein kinase II and (alpha)-actinin*. J Neurosci, 2001. 21(2): p. 423-33.
32. Kanamura, S. and J. Watanabe, *Cell biology of cytochrome P-450 in the liver*. Int Rev Cytol, 2000. 198: p. 109-52.
- 15 33. Ali, A., et al., *Design and synthesis of novel antibacterial agents with inhibitory activity against DNA polymerase III*. Bioorg Med Chem Lett, 2001. 11(16): p. 2185-8.
34. Weinberg, A.D., A.T. Vella, and M. Croft, *OX-40: life beyond the effector T cell stage*. Semin Immunol, 1998. 10(6): p. 471-80.
35. Kobayashi, K., et al., *Distribution and partial characterisation of IgG Fc binding protein in various mucin producing cells and body fluids*. Gut, 2002. 51(2): p. 169-76.
- 20 36. Miki, H., et al., *All kinesin superfamily protein, KIF, genes in mouse and human*. Proc Natl Acad Sci U S A, 2001. 98(13): p. 7004-11.
37. Peterson, M.R., S.C. Hsu, and R.H. Scheller, *A mammalian homologue of SLY1, a yeast gene required for transport from endoplasmic reticulum to Golgi*. Gene, 1996. 169(2): p. 293-4.
- 25 38. Penning, T.M., et al., *Generation of reactive oxygen species during the enzymatic oxidation of polycyclic aromatic hydrocarbon trans-dihydrodiols catalyzed by dihydrodiol dehydrogenase*. Chem Res Toxicol, 1996. 9(1): p. 84-92.
39. Coyle, P., et al., *Metallothionein: the multipurpose protein*. Cell Mol Life Sci, 2002. 59(4): p. 627-47.
- 30 40. Kang, S.W., et al., *Mammalian peroxiredoxin isoforms can reduce hydrogen peroxide generated in response to growth factors and tumor necrosis factor-alpha*. J Biol Chem, 1998. 273(11): p. 6297-302.
41. Banki, K., et al., *Glutathione levels and sensitivity to apoptosis are regulated by changes in transaldolase expression*. J Biol Chem, 1996. 271(51): p. 32994-3001.
- 35 42. Stolz, A., et al., *Molecular structure of rat hepatic 3 alpha-hydroxysteroid dehydrogenase. A member of the oxidoreductase gene family*. J Biol Chem, 1991. 266(23): p. 15253-7.

43. Methot, N., et al., *The human homologue of the yeast Prt1 protein is an integral part of the eukaryotic initiation factor 3 complex and interacts with p170*. J Biol Chem, 1997. 272(2): p. 1110-6.
- 5 44. Jesenberger, V. and S. Jentsch, *Deadly encounter: ubiquitin meets apoptosis*. Nat Rev Mol Cell Biol, 2002. 3(2): p. 112-21.
45. Singh, O.P., *Functional diversity of hnRNP proteins*. Indian J Biochem Biophys, 2001. 38(3): p. 129-34.
46. Huang, F., M. Wagner, and M.A. Siddiqui, *Structure, expression, and functional characterization of the mouse CLP- 1 gene*. Gene, 2002. 292(1-2): p. 245-59.
- 10 47. Tanaka, T., et al., *cDNA cloning and expression of rat homeobox gene, Hex, and functional characterization of the protein*. Biochem J, 1999. 339(Pt 1): p. 111-7.
48. Yu, F.X., et al., *Effects of thymosin beta 4 and thymosin beta 10 on actin structures in living cells*. Cell Motil Cytoskeleton, 1994. 27(1): p. 13-25.
- 15 49. Ventura-Holman, T., et al., *The murine fem1 gene family: homologs of the Caenorhabditis elegans sex- determination protein FEM-1*. Genomics, 1998. 54(2): p. 221-30.
50. Anderson, G.R. and D.L. Stoler, *Anoxia, wound healing, VL30 elements, and the molecular basis of malignant conversion*. Bioessays, 1993. 15(4): p. 265-72.
- 20 51. Anundi, I., et al., *Fructose prevents hypoxic cell death in liver*. Am J Physiol, 1987. 253(3 Pt 1): p. G390-6.
52. Kane, A.B., et al., *ATP depletion and loss of cell integrity in anoxic hepatocytes and silica-treated P388D1 macrophages*. Am J Physiol, 1985. 249(3 Pt 1): p. C256-66.
- 25 53. Pastorino, J.G., et al., *Cyclosporin and carnitine prevent the anoxic death of cultured hepatocytes by inhibiting the mitochondrial permeability transition*. J Biol Chem, 1993. 268(19): p. 13791-8.
54. Kaas GEN, J.M., Orrenius S., *Cyclosporine A protects hepatocytes against prooxidant-induced killing*. Biochem Pharmacol, 1992. 44: p. 1995-2003.
55. Schulte-Frohlinde D, S.C., *Radiolysis of DNA and model systems in the presence of oxygen*, in *Oxidative Stress*. 1985: Orlando. p. 11-40.
- 30 56. Tribble, D.L., T.Y. Aw, and D.P. Jones, *The pathophysiological significance of lipid peroxidation in oxidative cell injury*. Hepatology, 1987. 7(2): p. 377-86.
57. Bisgaard, H.C. and S.S. Thorgeirsson, *Hepatic regeneration. The role of regeneration in pathogenesis of chronic liver diseases*. Clin Lab Med, 1996. 16(2): p. 325-39.

35

#### OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the

scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.